Continuing to Explore New Frontiers:
Research to Integrate Administrative Data into the American Community Survey

ACS Webinar
May 13, 2020

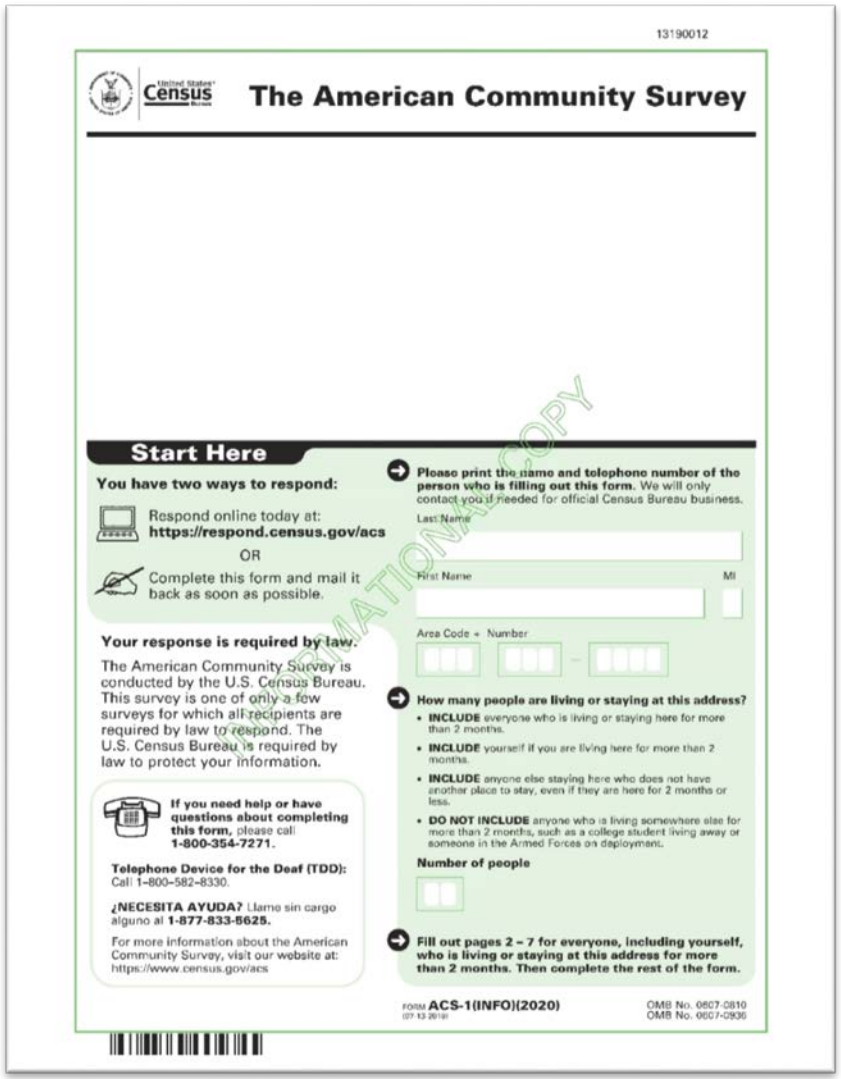
United States Census Bureau

This presentation is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical, methodological, operational, or technical issues are those of the presenters and not necessarily those of the U.S. Census Bureau.

American Community Survey

The American Community Survey is on the leading edge of survey design, continuous improvement, and data quality

- The nation's **most current, reliable, and accessible data** source for local statistics on critical planning topics such as age, children, veterans, commuting, education, income, and employment
- Surveys **3.5 million** addresses and informs over **$675 billion** of Federal government spending each year
- Covers **40+ topics**, supports over **300** evidence-based Federal government uses, and produces **11 billion** estimates each year
- Three key annual data releases:
  - 1-year Estimates (for large populations)
  - 1-year Supplemental Estimates (for small populations)
  - 5-year Estimates (for very small populations)



13190012

Census

The American Community Survey

Start Here

You have two ways to respond:

Respond online today at:
https://respond.census.gov/acs

OR

Complete this form and mail it back as soon as possible.

Your response is required by law.

The American Community Survey is conducted by the U.S. Census Bureau. This survey is one of only a few surveys for which all recipients are required by law to respond. The U.S. Census Bureau is required by law to protect your information.

If you need help or have questions about completing this form, please call **1-800-354-7271.**

Telephone Device for the Deaf (TDD): Call 1-800-582-8330.

¿NECESITA AYUDA? Llame sin cargo alguno al **1-877-833-5625.**

For more information about the American Community Survey, visit our website at: https://www.census.gov/acs

Please print the name and telephone number of the person who is filling out this form. We will only contact you if needed for official Census Bureau business.

Last Name

First Name MI

Area Code + Number

How many people are living or staying at this address?

- **INCLUDE** everyone who is living or staying here for more than 2 months.
- **INCLUDE** yourself if you are living here for more than 2 months.
- **INCLUDE** anyone else staying here who does not have another place to stay, even if they are here for 2 months or less.
- **DO NOT INCLUDE** anyone who is living somewhere else for more than 2 months, such as a college student living away or someone in the Armed Forces on deployment.

Number of people

Fill out pages 2 – 7 for everyone, including yourself, who is living or staying at this address for more than 2 months. Then complete the rest of the form.

FORM **ACS-1(INFO)(2020)**

OMB No. 0607-0810
OMB No. 0607-0936

United States Census Bureau

ACS Content Overview

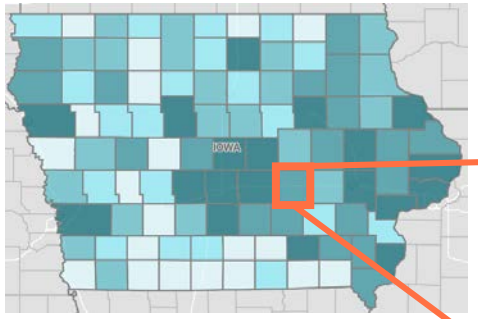
POPULATION		HOUSING
SOCIAL Ancestry Citizenship Disability Educational Attainment Fertility Grandparents Language Marital Status Migration School Enrollment Veterans	**DEMOGRAPHIC** Age Hispanic Origin Race Relationship Sex **ECONOMIC** Class of Worker Commuting Employment Status Food Stamps (SNAP) Health Insurance Hours/Week, Weeks/Year Income Industry & Occupation	Computer & Internet Use Costs (Mortgage, Rent, Taxes, Insurance) Heating Fuel Home Value Occupancy Plumbing/Kitchen Facilities Structure Tenure (Own/Rent) Utilities Vehicles Year Built/ Year Moved In

United States™ Census Bureau

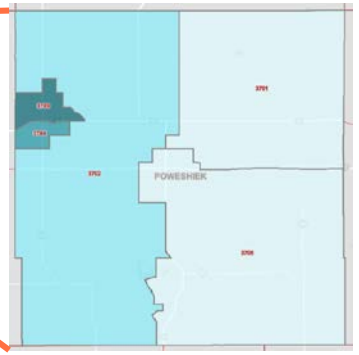
ACS Geographic Concepts

STATE

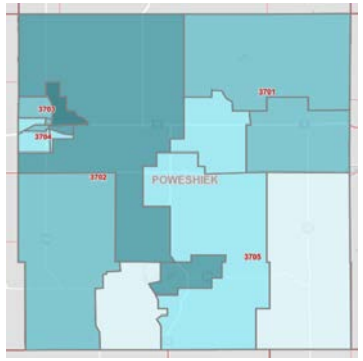
COUNTY/PLACE



CENSUS TRACT



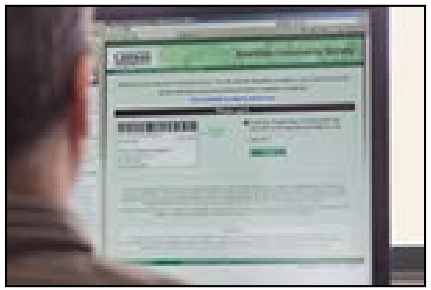
BLOCK GROUP



United States Census Bureau

ACS Data Collection Process

Online



Mail

Self-Response
(about 6 weeks)

(Between 17 and 24 days)

Personal Visit[1]

Nonresponse Follow-up
(about 4 weeks)

United States Census Bureau

[1] The Personal Visit is conducted via Computer Assisted Personal Interviewing (CAPI).

Why and How to Use Administrative Data?

Increase data quality by filling in missing responses and using administrative data to evaluate and enrich survey data

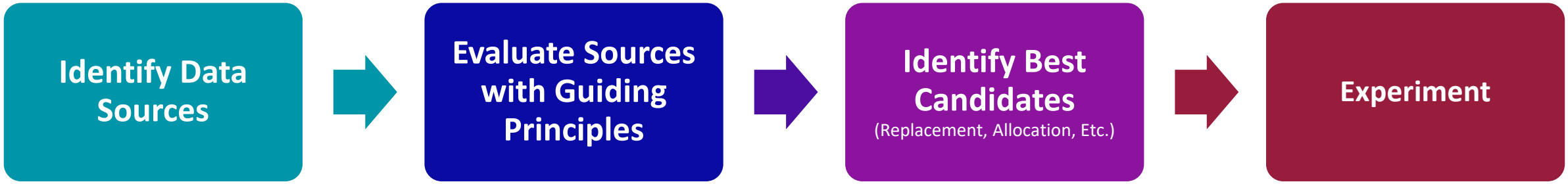
Save time and improve respondent experience by reducing the number of questions asked on the ACS

Provide cost savings by identifying vacant housing units and reducing the need for follow-up visits

Mandated by Title 13 of the U.S. Code to use already available information

United States™ Census Bureau

Charting the Course



United States Census Bureau

Charting the Course

Identify Data Sources

- **Federal Data**
  (e.g., Internal Revenue Service, Social Security Administration)
- **State and Local Data**
  (e.g., Temporary Assistance for Needy Families, Supplemental Nutrition Assistance Program)
- **Third-Party Data**
  (e.g., Black Knight property and tax foreclosure, VSGI consumer households)

→ **Evaluate Sources with Guiding Principles**

→ **Identify Best Candidates**
(Replacement, Allocation, Etc.)

→ **Experiment**

United States Census Bureau

Charting the Course

Identify Data Sources

- **Federal Data** (e.g., Internal Revenue Service, Social Security Administration)
- **State and Local Data** (e.g., Temporary Assistance for Needy Families, Supplemental Nutrition Assistance Program)
- **Third-Party Data** (e.g., Black Knight property and tax foreclosure, VSGI consumer households)

Evaluate Sources with Guiding Principles

- Authorization
- Availability
- Conceptual Alignment
- Coverage
- Data Source
- Disclosure Avoidance
- Impacts on Estimates
- Intended Use
- Population Universe
- Quality
- Reliability
- Temporal Alignment

Identify Best Candidates
(Replacement, Allocation, Etc.)

Experiment

United States™ Census Bureau

Charting the Course

Identify Data Sources

Federal Data
(e.g., Internal Revenue Service, Social Security Administration)

State and Local Data
(e.g., Temporary Assistance for Needy Families, Supplemental Nutrition Assistance Program)

Third-Party Data
(e.g., Black Knight property and tax foreclosure, VSGI consumer households)

Evaluate Sources with Guiding Principles

- Authorization
- Availability
- Conceptual Alignment
- Coverage
- Data Source
- Disclosure Avoidance
- Impacts on Estimates
- Intended Use
- Population Universe
- Quality
- Reliability
- Temporal Alignment

Identify Best Candidates

(Replacement, Allocation, Etc.)

Most Promising:
Property Value
Real Estate Tax
Year Built
Acreage

Less Promising:
Part of a Condominium
Phone Service
Facilities (Kitchen/Bathroom)
Fuel Type
Tenure
Have a Mortgage
First/Secondary Mortgage Payment
Number of (Bed)Rooms
Agricultural Sales

Experiment

United States™ Census Bureau

Charting the Course

Identify Data Sources

Federal Data
(e.g., Internal Revenue Service, Social Security Administration)

State and Local Data
(e.g., Temporary Assistance for Needy Families, Supplemental Nutrition Assistance Program)

Third-Party Data
(e.g., Black Knight property and tax foreclosure, VSGI consumer households)

Evaluate Sources with Guiding Principles

- Authorization
- Availability
- Conceptual Alignment
- Coverage
- Data Source
- Disclosure Avoidance
- Impacts on Estimates
- Intended Use
- Population Universe
- Quality
- Reliability
- Temporal Alignment

Identify Best Candidates

(Replacement, Allocation, Etc.)

Most Promising:
Property Value
Real Estate Tax
Year Built
Acreage

Less Promising:
Part of a Condominium
Phone Service
Facilities (Kitchen/Bathroom)
Fuel Type
Tenure
Have a Mortgage
First/Secondary Mortgage Payment
Number of (Bed)Rooms
Agricultural Sales

Experiment

Predicting Vacant Housing Units to reduce need for follow-up visits

Replace survey responses with administrative data to reduce the number of questions asked of respondents

Fill in missing values to improve data quality

Enrich ACS data through linkages to administrative data to expand topics covered

Use administrative records to improve measurement of income

United States™ Census Bureau

Today's Presentations

Predicting Vacant Housing Units in the ACS

Andrew Keller | Andrew.D.Keller@census.gov

Analyzing Differences between Survey Responses and Administrative Data for Property Value

R. Chase Sawyer | Robert.C.Sawyer@census.gov

Using Alternative Data Sources to Fill-in Missing Values for Demographic Characteristics in the ACS

Sandra L. Clark | Sandra.L.Clark@census.gov

Linking ACS and IRS Data to Assess College Attendance and Completion by Family Income

Leah Clark | Leah.R.Clark@census.gov

United States™ Census Bureau

Predicting Vacant Housing Units in the American Community Survey

Andrew Keller
Decennial Statistical Studies Division

United States Census Bureau

Outline

- Background
- Research Objective and Methods
- Simulation
- Discussion

United States Census Bureau

Background

- 2020 Census
  - Can we use administrative records (AR) to inform Nonresponse Followup (NRFU) operation?
  - Use AR to model occupied, vacant, delete statuses of NRFU universe
  - Modify contact strategy where we have high confidence in address status
  - Conduct one visit to units with high confidence of final status via models
- Apply same concept to predict units with high confidence of vacancy in American Community Survey (ACS)
  - Conceptual Difference: 2020 Census determination is point-in-time
- Use probability to inform contact or sampling strategy

United States™ Census Bureau

Model

- Logistic Model: Dependent Variable is vacancy outcome status
- Use Previous Year(s) of ACS to form training data
  - Administrative Records data of the same vintage
  - Operational Data
  - Address-Level information from Master Address File
  - Block Group-level information from ACS Planning Database
- Apply parameter estimates to current vintage of ACS data
- Predicted probability of vacancy for every ACS unit in mailable Computer Assisted Personal Interview (CAPI) universe

United States™ Census Bureau

Data

- Administrative Records data
  - Aggregated public information purchased by Census Bureau consisting of local tax, deed, and mortgage information
  - Using information concerning land use, absence of owner at address, ownership rights on the unit
  - Third-Party AR data providing information about persons at addresses
  - National Change of Address information from United States Postal Service (USPS)

Data

- Administrative Records data
  - Aggregated public information purchased by Census Bureau consisting of local tax, deed, and mortgage information
  - Using information concerning land use, absence of owner at address, ownership rights on the unit
  - Third-Party AR data providing information about persons at addresses
  - National Change of Address information from United States Postal Service (USPS)
- Operational Data
  - Mailing operations (undeliverable as addressed from USPS)
  - Indication of vacancy from internet response

Data

- Administrative Records data
  - Aggregated public information purchased by Census Bureau consisting of local tax, deed, and mortgage information
  - Using information concerning land use, absence of owner at address, ownership rights on the unit
  - Third-Party AR data providing information about persons at addresses
  - National Change of Address information from United States Postal Service (USPS)
- Operational Data
  - Mailing operations (undeliverable as addressed from USPS)
  - Indication of vacancy from internet response
- Address-Level information from Master Address File
  - Delivery Sequence File status (Residential, Commercial, Excluded from Delivery Statistics)
  - Housing Unit Type (Multi, Single, Trailer, Other)
  - Delivery Point Type

United States™ Census Bureau

Data

- Administrative Records data
  - Aggregated public information purchased by Census Bureau consisting of local tax, deed, and mortgage information
  - Using information concerning land use, absence of owner at address, ownership rights on the unit
  - Third-Party AR data providing information about persons at addresses
  - National Change of Address information from United States Postal Service (USPS)
- Operational Data
  - Mailing operations (undeliverable as addressed from USPS)
  - Indication of vacancy from internet response
- Address-Level information from Master Address File
  - Delivery Sequence File status (Residential, Commercial, Excluded from Delivery Statistics)
  - Housing Unit Type (Multi, Single, Trailer, Other)
  - Delivery Point Type
- Block Group-level information from ACS Planning Database
  - Poverty, Rental, Other Language rates, Hispanic

United States™ Census Bureau

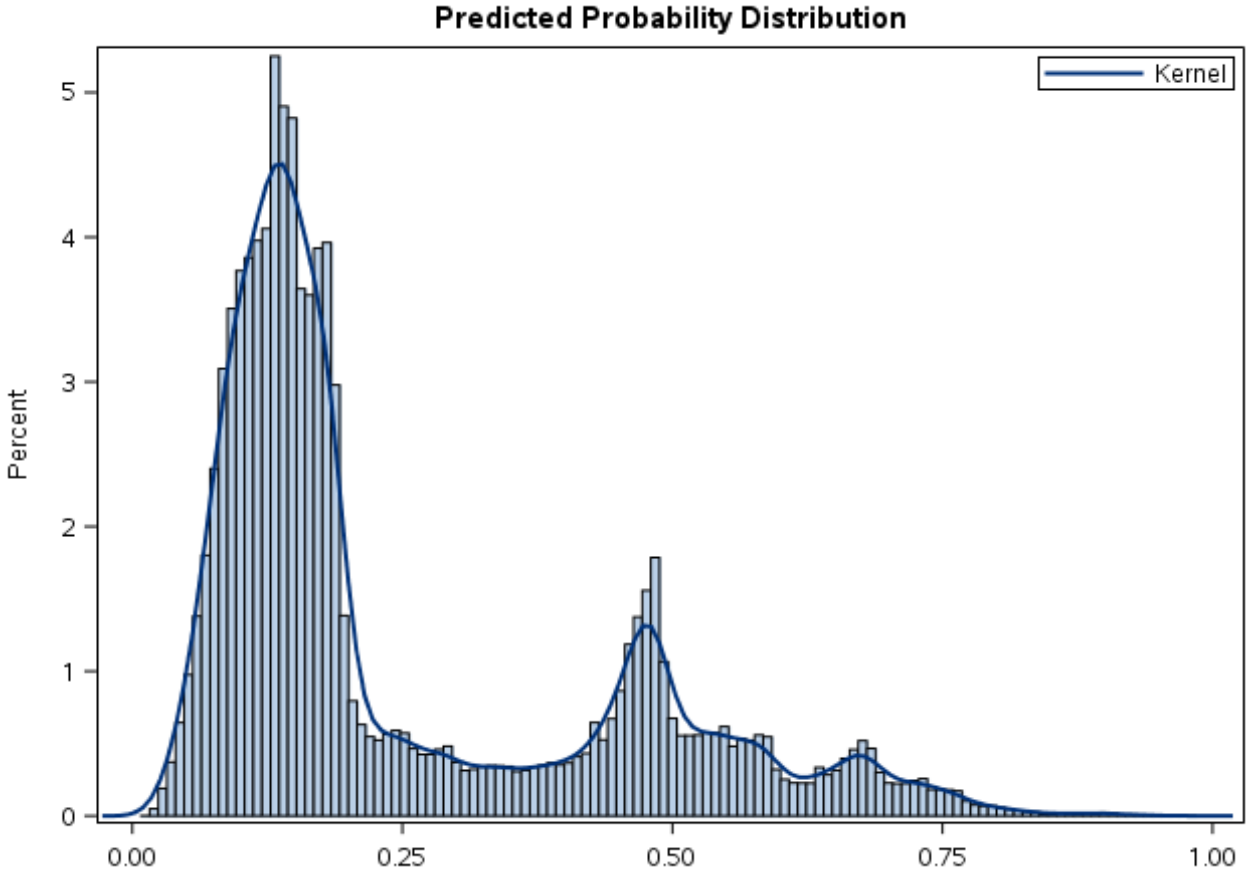
Simulation Setup

- Take mailable CAPI universe cases from 2016, 2017 ACS universe
- Fit model on 2016 data
- Score model on 2017 data
- Sort predicted vacant probabilities from greatest to least
- Iterate over top percentages by picking a threshold. (i.e. – top 10% or 5% of predicted vacant probabilities)
- See how many of those were vacant in 2017 (About 25% of universe is vacant)

United States™ Census Bureau

Simulation Results – Predictions

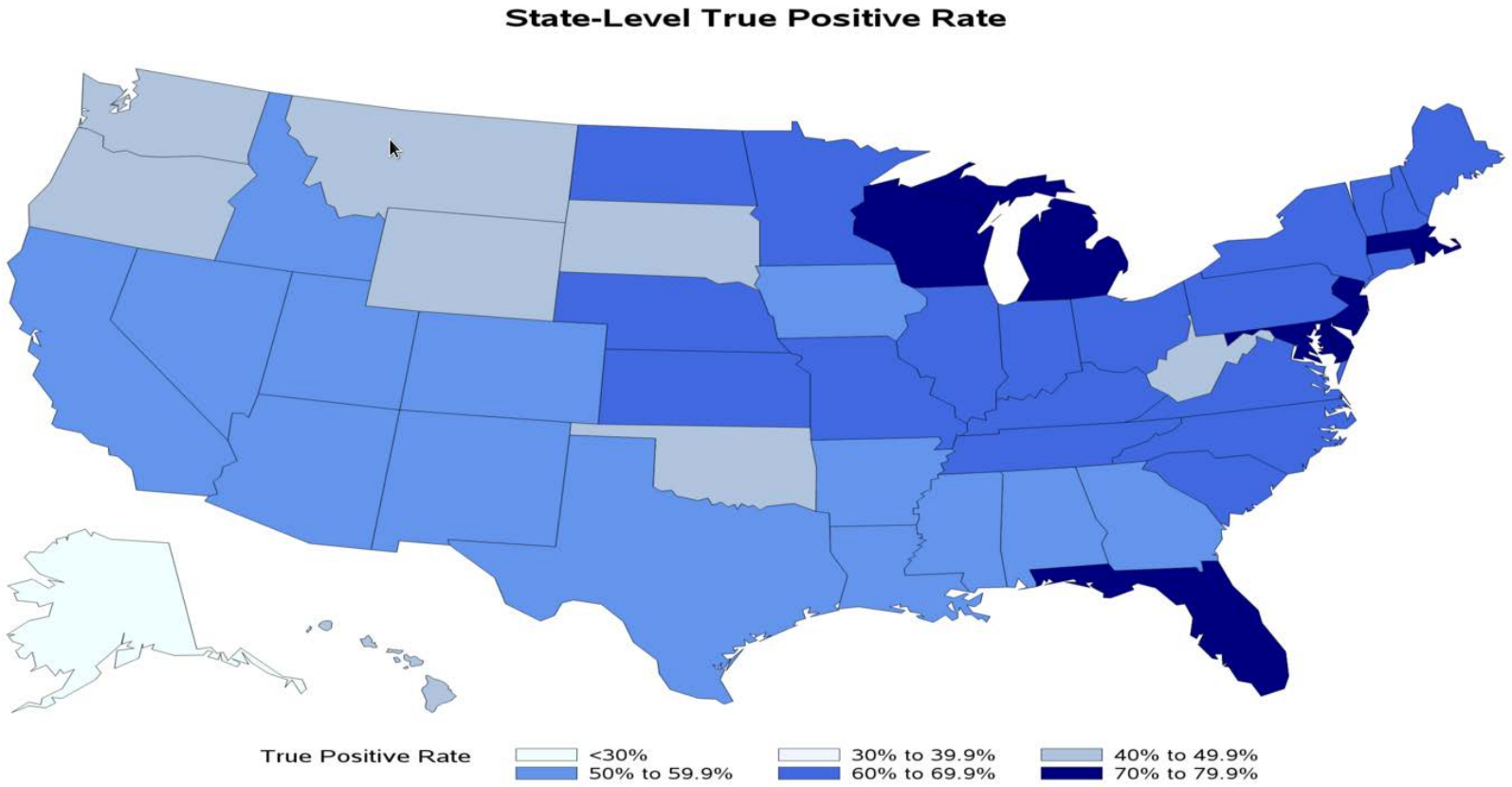
- 1) Can we design a model to reasonably predict vacant cases?
- 2) What threshold should be used to perform treatment?



United States Census Bureau

Simulation Results – State-Level Predictions

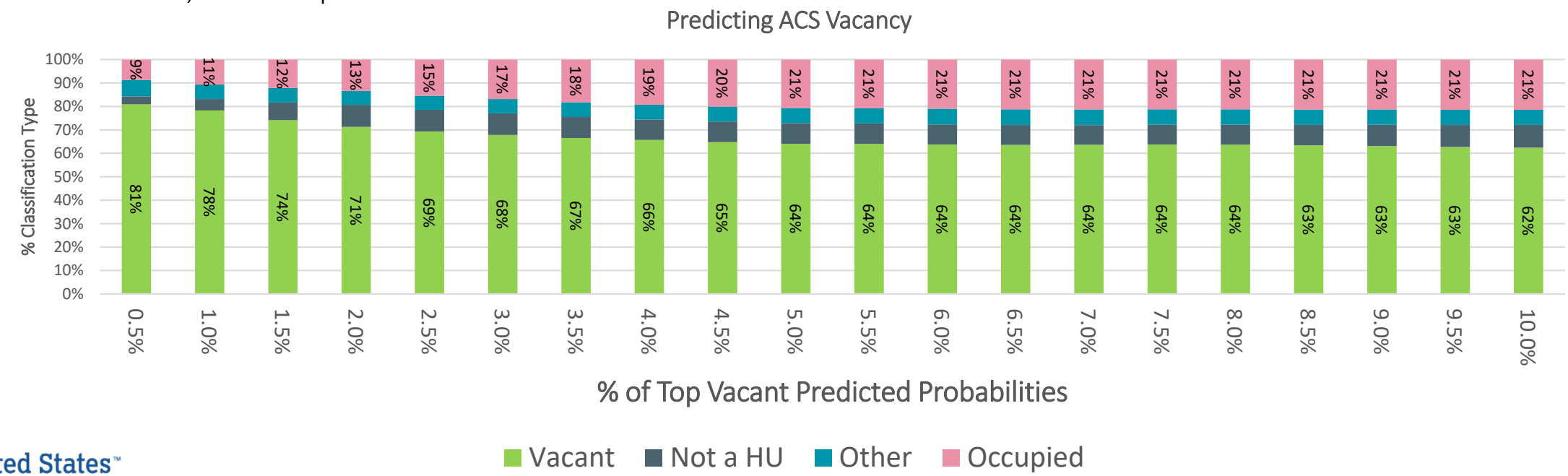
- For Top 10% of vacant predicted probabilities, what is the rate at which the vacancy prediction is correct?



United States Census Bureau

Simulation Results – Status

- Look at distribution of outcome
- Example: Top 5.0% of vacant predicted probability among mailable CAPI cases:
  - 64% vacant, 21% occupied



United States Census Bureau

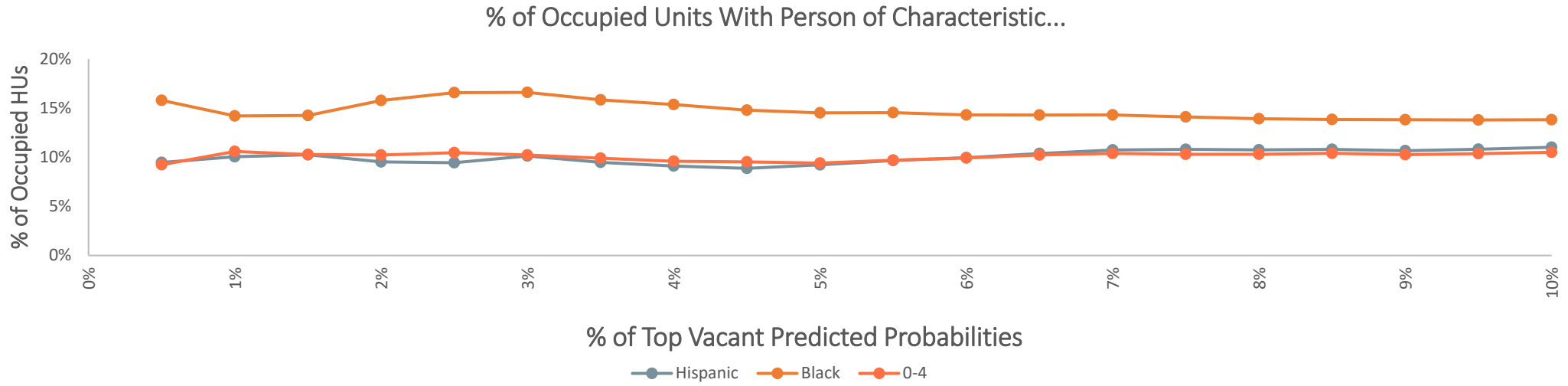
Simulation Results – State-Level Treatment

- The universe of mailable CAPI cases has a particular state-level distribution.
- Depending on cost-benefit, we pick a threshold of predicted vacant probabilities to treat.
- Regardless of the top threshold selected, that state-level distribution of the top threshold will not be the same.

United States™ Census Bureau

Simulation Results – People in HUs

- Among occupied mailable CAPI cases:
  - 25% have someone who is Hispanic
  - 21% have someone who is Black
  - 15% have someone who is age 0-4
- For the top vacant predicted probabilities, we identify occupied units that have proportionally fewer Hispanic, Black, 0-4 persons



United States Census Bureau

Analysis of False Positives

- Study top vacant predicted probabilities with non-vacant outcome
- Develop understanding where we might be more sensitive to calling it vacant
  - Use decision tree to create business rules
  - Develop rules where non-vacant outcomes occur in greatest amount
- Example Rule
  - Take top 10% of predicted cases
  - No Land Use Indicated on Administrative Records
  - Not on Delivery Sequence File of previous fall
    - 15% of universe
    - 52% were not vacant

United States™ Census Bureau

Conclusions

- Modeling vacant units in the ACS universe can be completed using a combination of address-level, ACS operational, geographic, and administrative records information.
- Cost-benefit analysis will help determine the threshold for using the best predictions.
  - Generally, cases within the chosen thresholds contain relatively fewer members of the hard-to-count groups.
- We observe differential true positive rates across states.
  - We develop business rules to identify false positive cases.
  - Not all false positive cases are occupied units – some are addresses without housing units.

United States™ Census Bureau

Discussion and Future Work

- Universe does not have to be mailable CAPI cases – can include all CAPI cases

- We can use ACS Contact History to update model.
  - Example – feed results from first contact into model, update probabilities with that information

- We can use predicted probabilities to alter contact strategy.
  - This is how the information is being applied for the 2020 Census.

- We can use predicted probabilities to change sampling rates.
  - Risk: Changing sampling rate for high probability vacant cases that are occupied inflates variances

United States™ Census Bureau

Analyzing Differences between Survey Responses and Administrative Data for Property Value

R. Chase Sawyer
American Community Survey Office

United States Census Bureau

Housing Administrative Record Simulation

- Research goals
  - Simulate use of administrative records (ARs) in ACS
  - Study impact on data products
  - Test feasibility of implementing methods
  - Learn effects on production process

United States® Census Bureau

Simulating this Design

- Used 2015 ACS Responses
- Direct substitution for the four most promising housing items:
  - Year built
  - Acreage
  - Real estate taxes

  } Property tax records from vendor CoreLogic

  - Property value

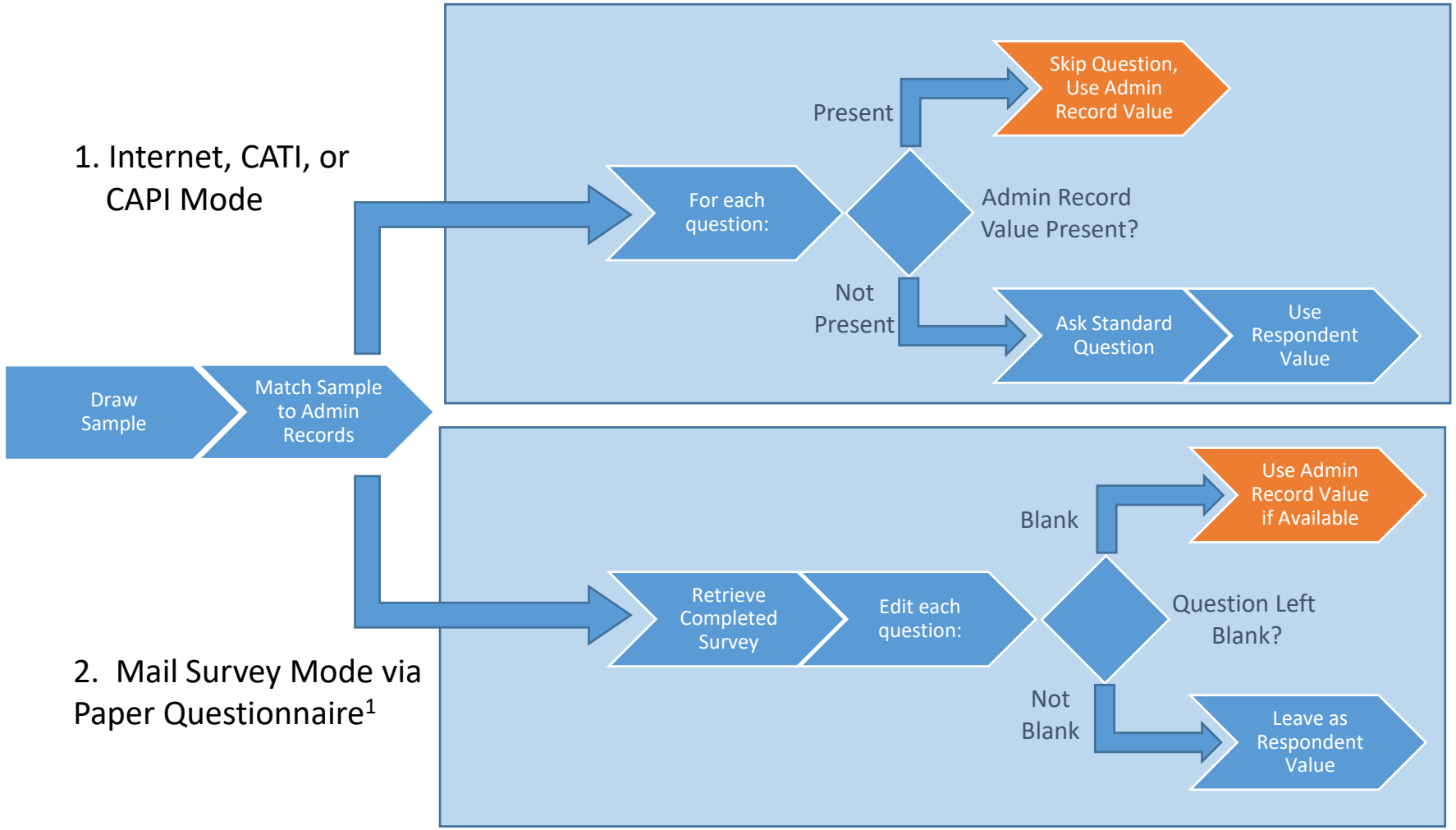
  } Modeled admin data from CoreLogic's Automated Valuation Model (AVM)
- Produced **"Simulated"** version to compare to **"Published"** 2015 ACS

United States® Census Bureau

Adaptive Design

1. Internet, CATI, or CAPI Mode

2. Mail Survey Mode via Paper Questionnaire[1]



1. It would not be feasible to have multiple versions of the mail form so we assume we would ask these questions of all mail respondents.

Limitations

- Simulated estimates, not a direct comparison
- Results may be confounded by linkage

United States® Census Bureau

Key Measures – Property Value

Key Measure	Published	Simulated	Percent Difference	MOE
Median property value	$194,500	$182,300	-6.3	0.1
Property value less than $10,000	1,045,716	875,020	-16.3	0.6
Property value $2,000,000 or more	555,865	407,895	-26.6	1.0

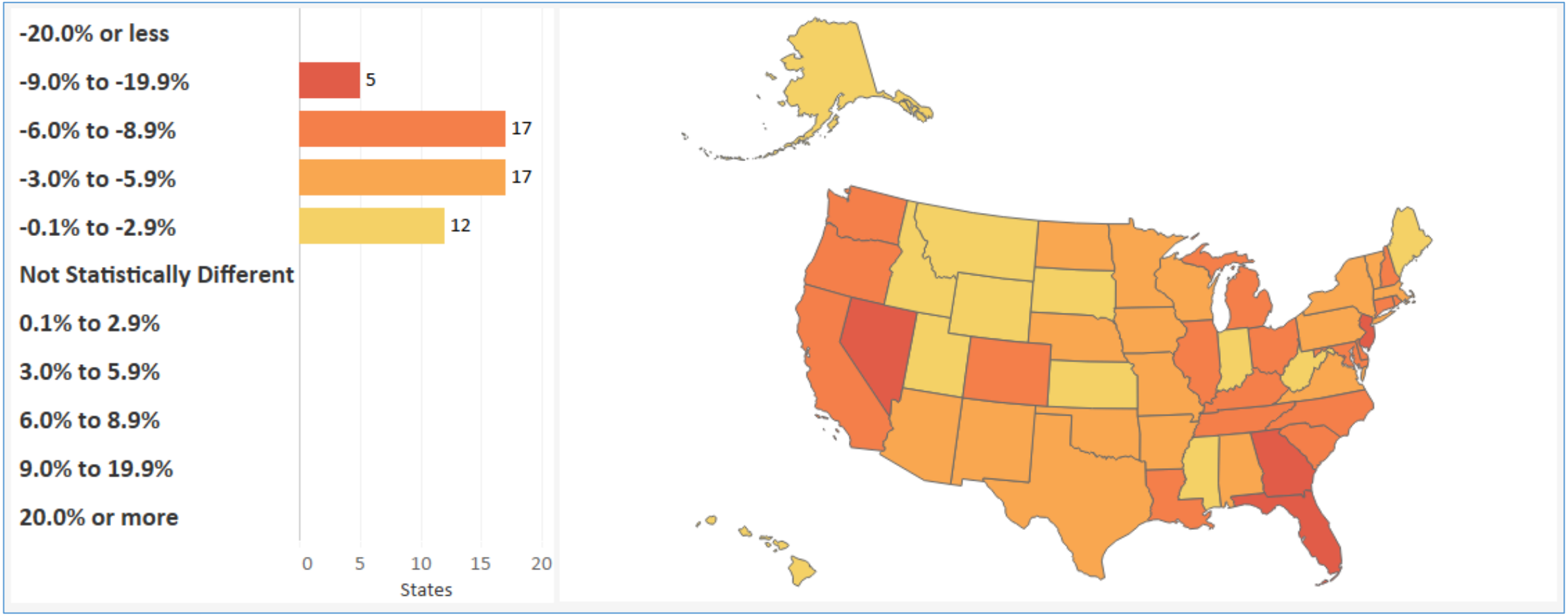
United States® Census Bureau

Key Measures – Property Value

Key Measure	Published	Simulated	Percent Difference	MOE
Median property value	**$194,500**	**$182,300**	**-6.3**	**0.1**
Property value less than $10,000	1,045,716	875,020	-16.3	0.6
Property value $2,000,000 or more	555,865	407,895	-26.6	1.0

United States® Census Bureau

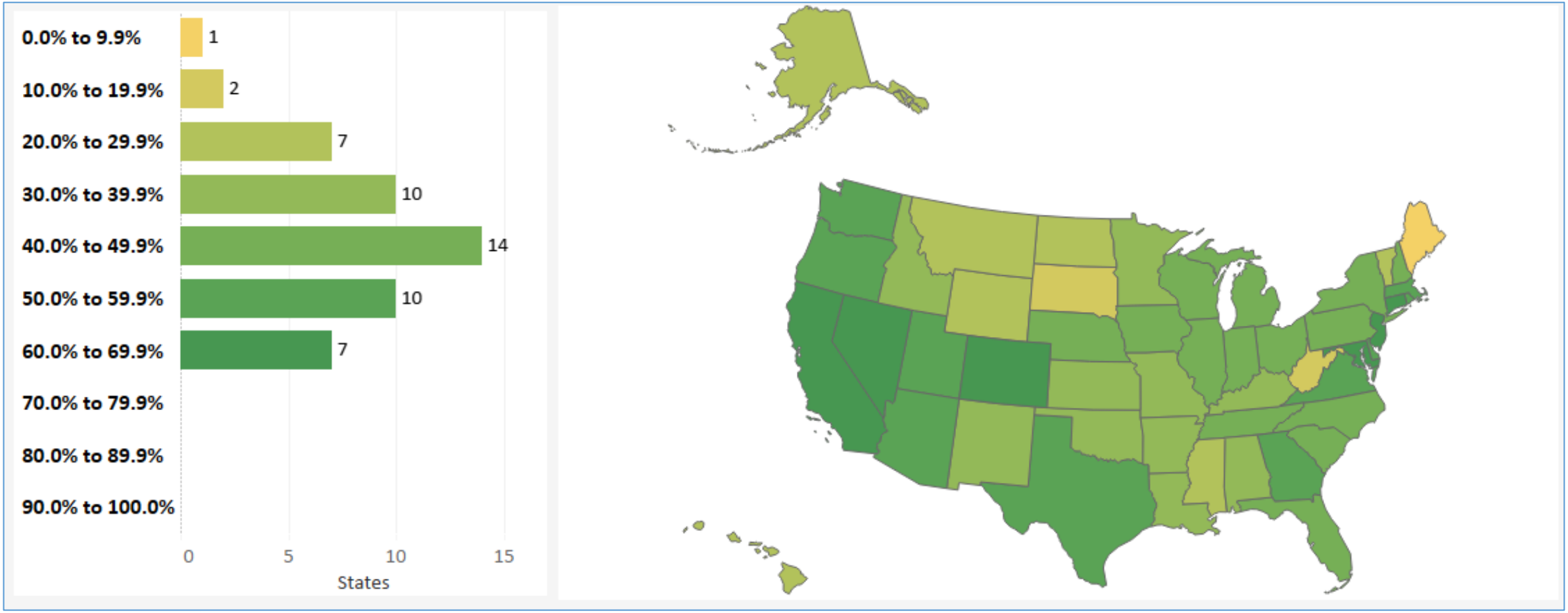
Percent Difference in Median Property Value: Simulated minus Published - State



Source: Housing Administrative Record Simulation (Clark et al, 2018)

United States® Census Bureau

Burden Reduction for Property Value Question - State



Source: Housing Administrative Record Simulation (Clark et al, 2018)

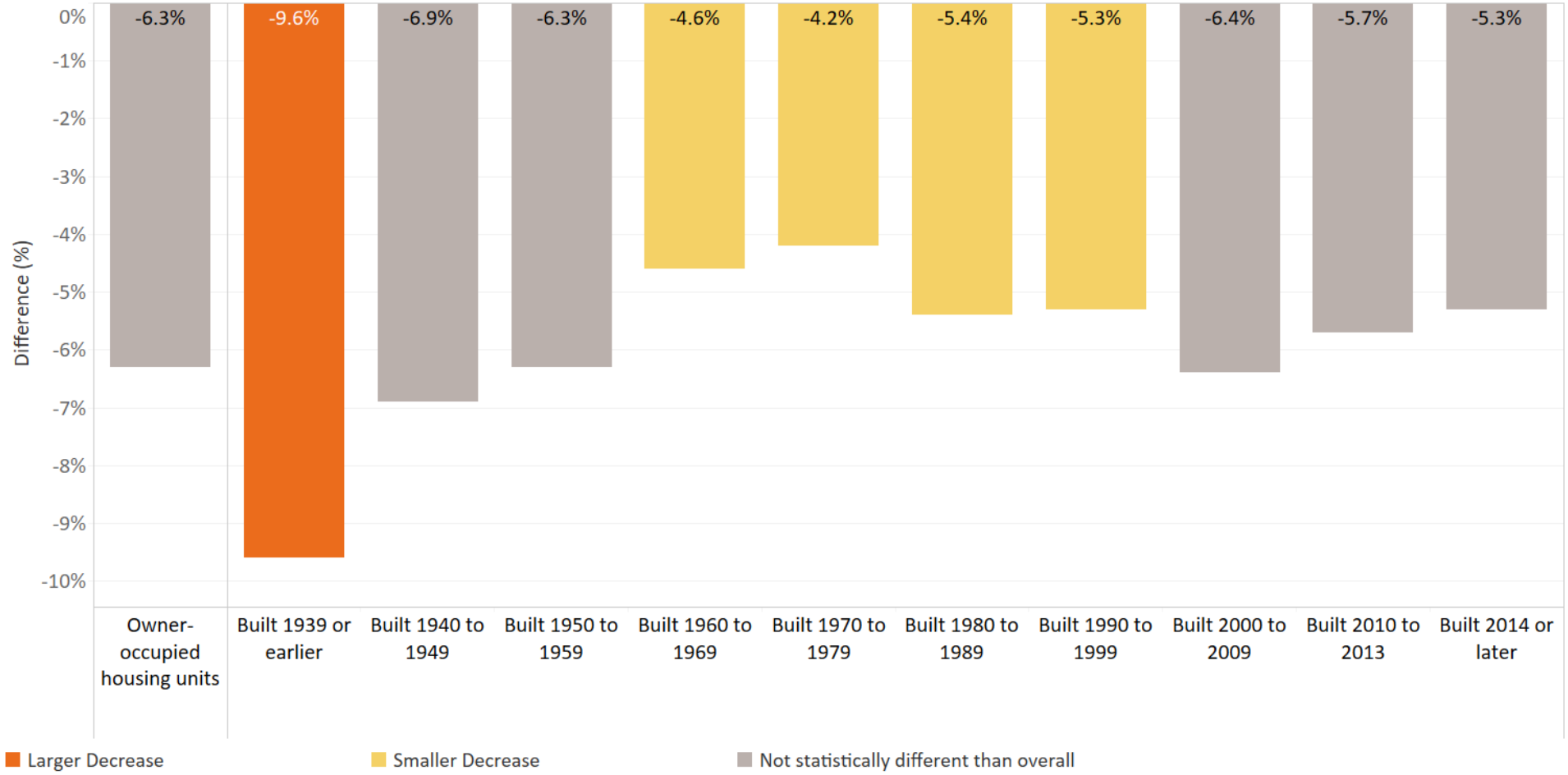
United States® Census Bureau

Additional data breakouts

- Median property value
  - By year built
  - By year moved in
  - By mortgage status
- Distribution of property value
  - Overall
  - By household income

United States® Census Bureau

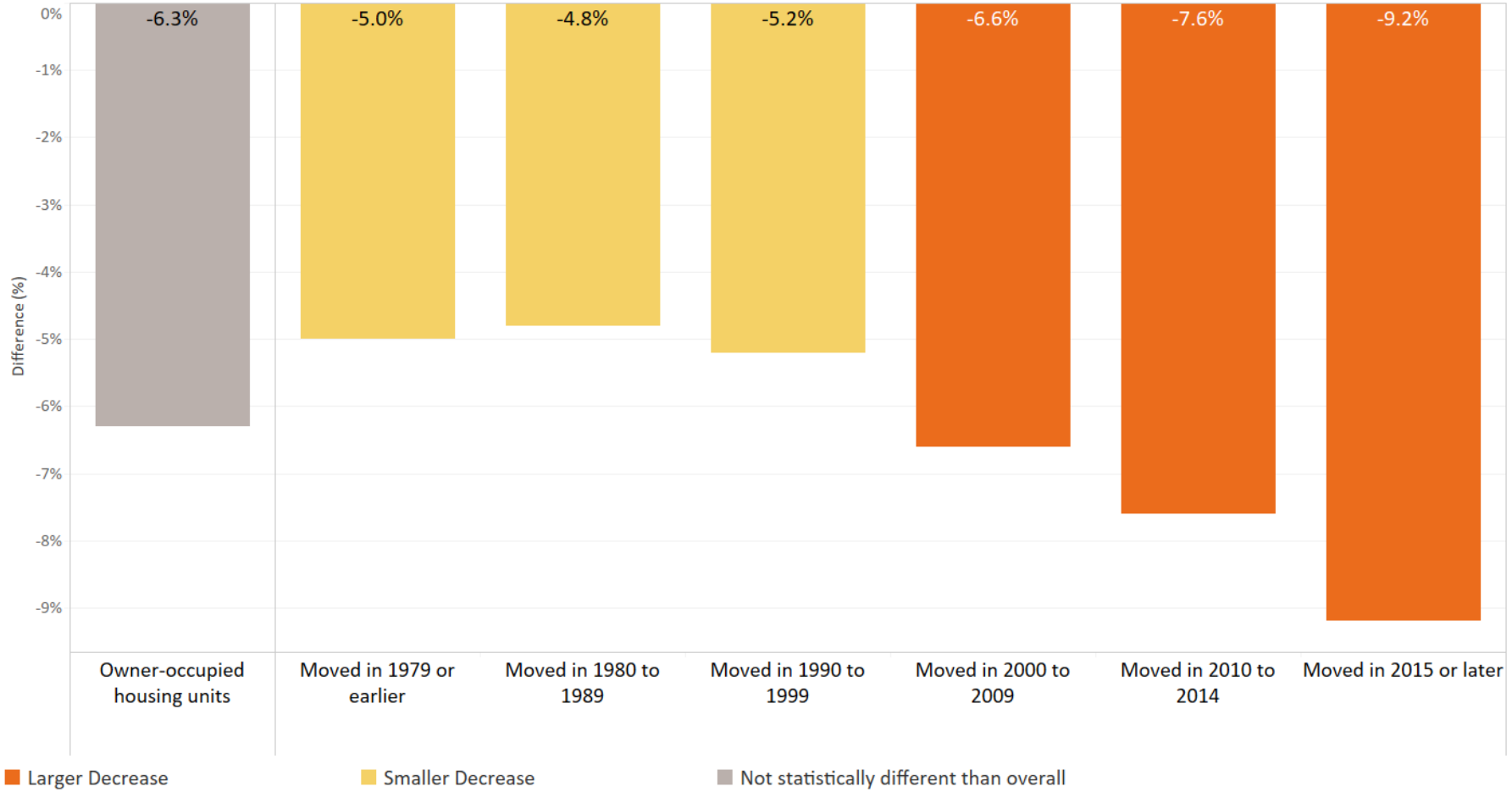
Differences in Median Property Value by Year Built



Source: Housing Administrative Record Simulation (Clark et al, 2018)

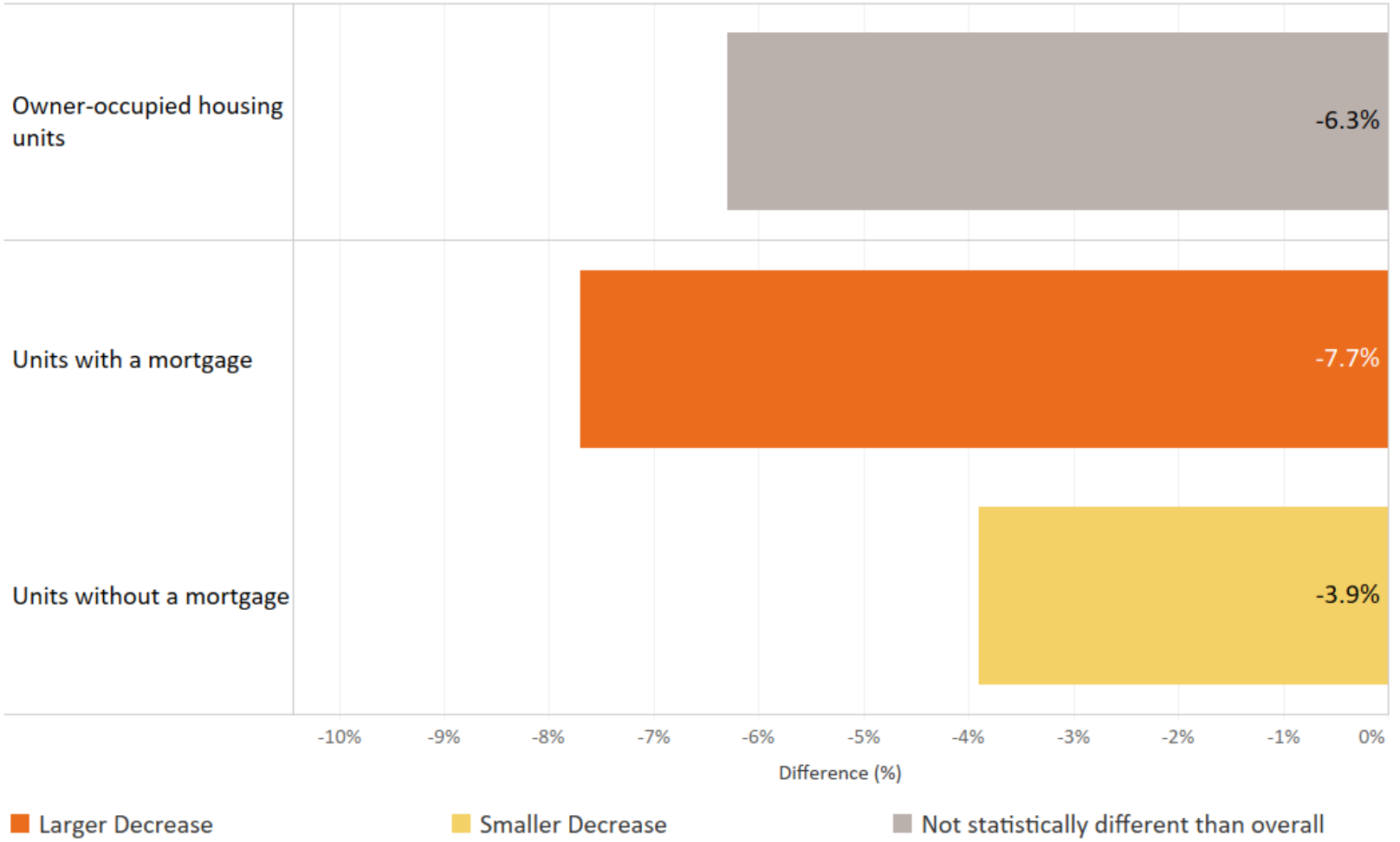
United States® Census Bureau

Differences in Median Property Value by Year Moved In



Source: Housing Administrative Record Simulation (Clark et al, 2018)

Differences in Median Value by Mortgage Status



Source: Housing Administrative Record Simulation (Clark et al, 2018)

Takeaways – Differences in Median Property Value

- Year built
  - No discernable pattern, units built before 1940 have largest decrease
- Year moved in
  - Units that have been moved into more recently have largest decrease
- Mortgage status
  - Decrease largest for units that have a mortgage

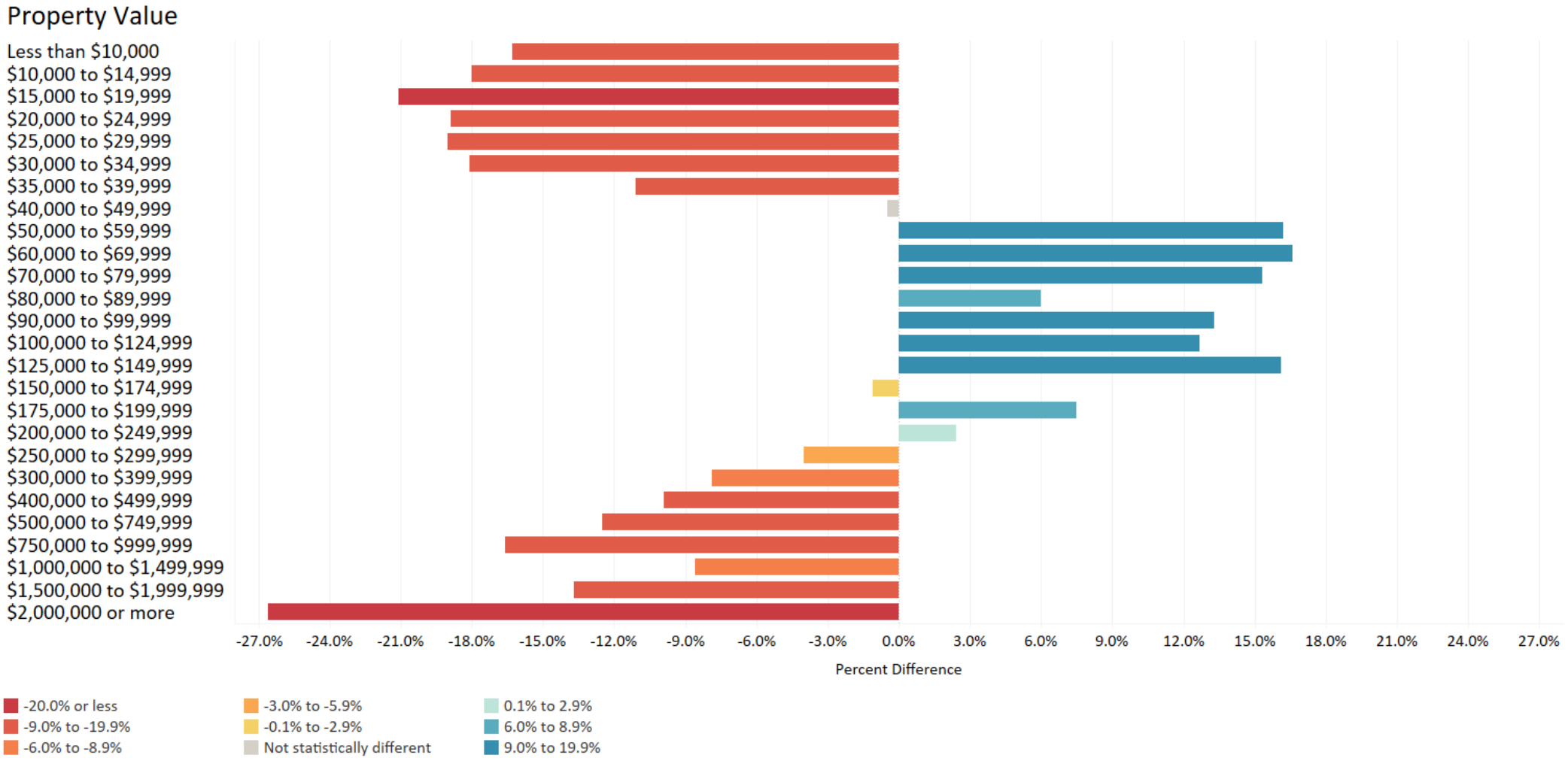
United States® Census Bureau

Key Measures – Property Value

Key Measure	Published	Simulated	Percent Difference	MOE
Median property value	$194,500	$182,300	-6.3	0.1
Property value less than $10,000	1,045,716	875,020	-16.3	0.6
Property value $2,000,000 or more	555,865	407,895	-26.6	1.0

United States® Census Bureau

Difference in the Number of Households – Property Value



Source: Housing Administrative Record Simulation (Clark et al, 2018)

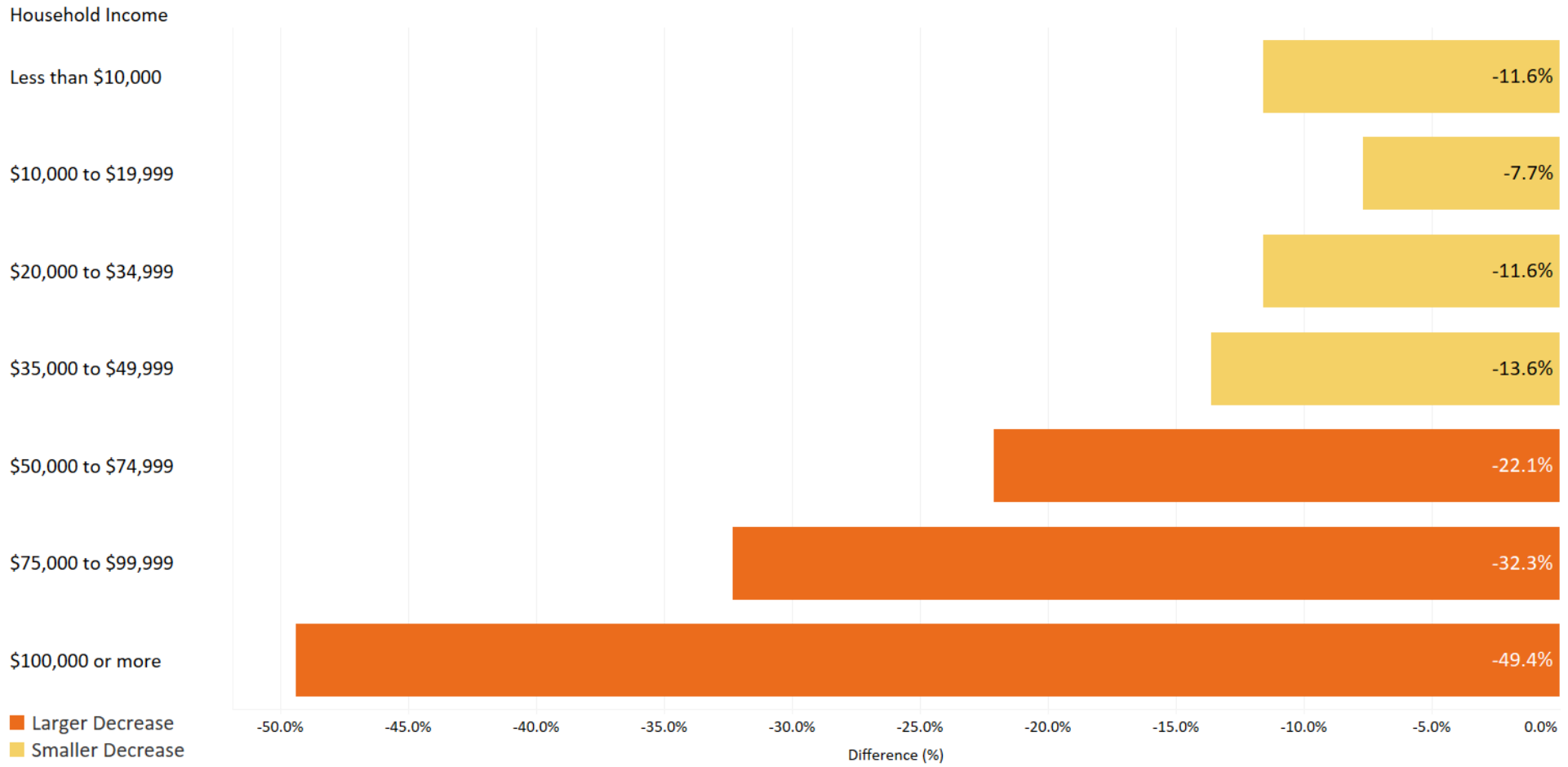
United States® Census Bureau

Key Measures - Value

Key Measure	Simulated	Published	Percent Difference	MOE
Median property value	$182,300	$194,500	-6.3	0.1
Property value less than $10,000	875,020	1,045,716	-16.3	0.6
Property value $2,000,000 or more	407,895	555,865	-26.6	1.0

United States® Census Bureau

Difference in the Number of Households with a Property Value of Less than $10,000 by household income



Source: Housing Administrative Record Simulation (Clark et al, 2018)

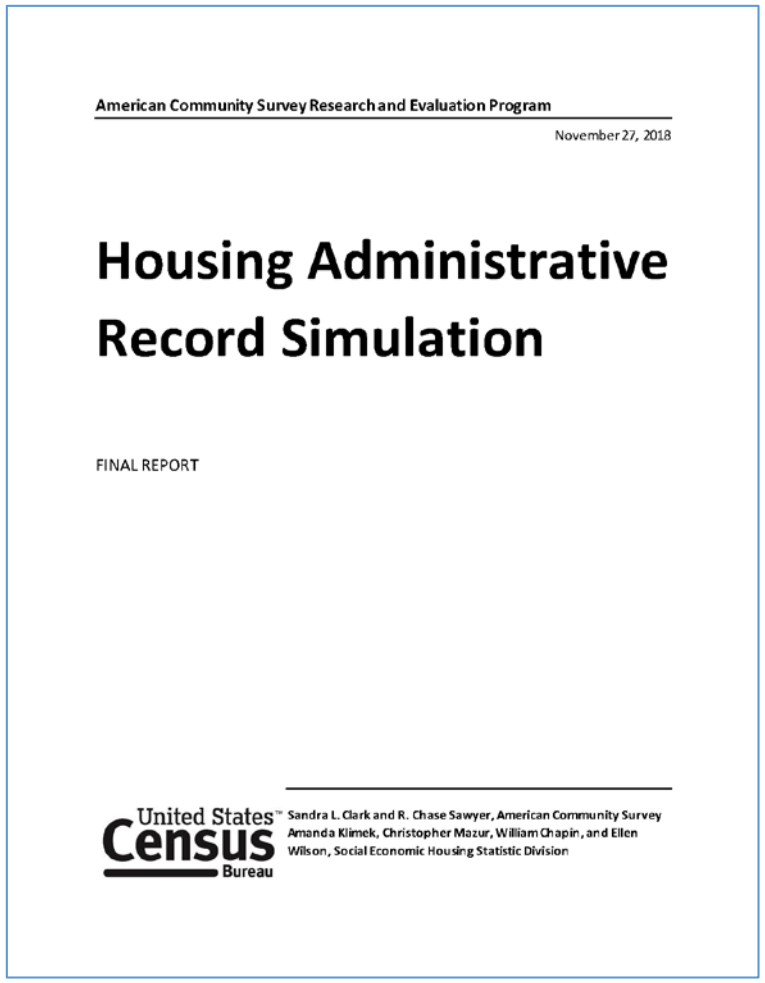
United States® Census Bureau

Next steps

- Compare ACS responses to new administrative record data
- Look specific at housing characteristics discussed today
- Possible modeling projects

United States® Census Bureau

Research Products



American Community Survey Research and Evaluation Program

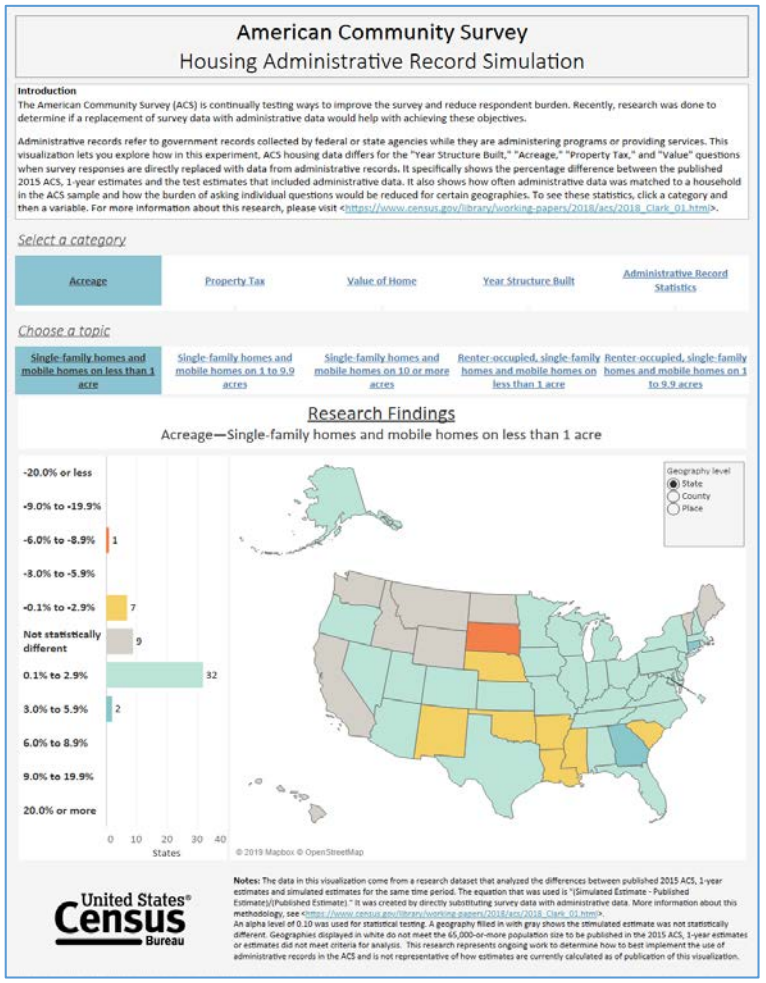
November 27, 2018

Housing Administrative Record Simulation

FINAL REPORT

United States Census Bureau

Sandra L. Clark and R. Chase Sawyer, American Community Survey
Amanda Klimek, Christopher Mazur, William Chapin, and Ellen Wilson, Social Economic Housing Statistic Division



American Community Survey
Housing Administrative Record Simulation

Introduction

The American Community Survey (ACS) is continually testing ways to improve the survey and reduce respondent burden. Recently, research was done to determine if a replacement of survey data with administrative data would help with achieving these objectives.

Administrative records refer to government records collected by federal or state agencies while they are administering programs or providing services. This visualization lets you explore how in this experiment, ACS housing data differs for the "Year Structure Built," "Acreage," "Property Tax," and "Value" questions when survey responses are directly replaced with data from administrative records. It specifically shows the percentage difference between the published 2015 ACS, 1-year estimates and the test estimates that included administrative data. It also shows how often administrative data was matched to a household in the ACS sample and how the burden of asking individual questions would be reduced for certain geographies. To see these statistics, click a category and then a variable. For more information about this research, please visit <https://www.census.gov/library/working-papers/2018/acs/2018_Clark_01.html>.

Select a category

Acreage | Property Tax | Value of Home | Year Structure Built | Administrative Record Statistics

Choose a topic

Single-family homes and mobile homes on less than 1 acre | Single-family homes and mobile homes on 1 to 9.9 acres | Single-family homes and mobile homes on 10 or more acres | Renter-occupied, single-family homes and mobile homes on less than 1 acre | Renter-occupied, single-family homes and mobile homes on 1 to 9.9 acres

Research Findings

Acreage—Single-family homes and mobile homes on less than 1 acre

United States Census Bureau

Notes: The data in this visualization come from a research dataset that analyzed the differences between published 2015 ACS, 1-year estimates and simulated estimates for the same time period. The equation that was used is "(Simulated Estimate - Published Estimate)/(Published Estimate)." It was created by directly substituting survey data with administrative data. More information about this methodology, see <https://www.census.gov/library/working-papers/2018/acs/2018_Clark_01.html>.
An alpha level of 0.10 was used for statistical testing. A geography filled in with gray shows the stimulated estimate was not statistically different. Geographies displayed in white do not meet the 65,000-or-more population size to be published in the 2015 ACS, 1-year estimates or estimates did not meet criteria for analysis. This research represents ongoing work to determine how to best implement the use of administrative records in the ACS and is not representative of how estimates are currently calculated as of publication of this visualization.

United States® Census Bureau

census.gov/library/working-papers/2018/acs/2018_Clark_01.html
census.gov/library/visualizations/interactive/housing-admin-record-simulation.html

Using Alternative Data Sources to Fill-in Missing Values for Demographic Characteristics in the ACS

Sandra L. Clark
American Community Survey Office

United States Census Bureau

Why and How to Use Administrative Data?

- Increase data quality by filling in missing responses and using administrative data to evaluate and enrich survey data
- Save time and improve respondent experience by reducing the number of questions asked on the ACS
- Provide cost savings by identifying vacant housing units and reducing the need for follow-up visits
- Mandated by Title 13 of the U.S. Code to use already available information

United States™ Census Bureau

Background

- AR more accurate than statistical approaches such as hotdeck imputation
- Research for 2020 Decennial Census
  - Hispanic Origin, Race, Age
  - High match rates between 2010 Census reported and AR - 90% or better
  - Differences in imputed 2010 Census and AR
    - Increase in Hispanic and race categories when replacing imputed cases
    - Census Imputes more older ages than younger when compared to AR
- Using AR for imputation relatively easy to implement

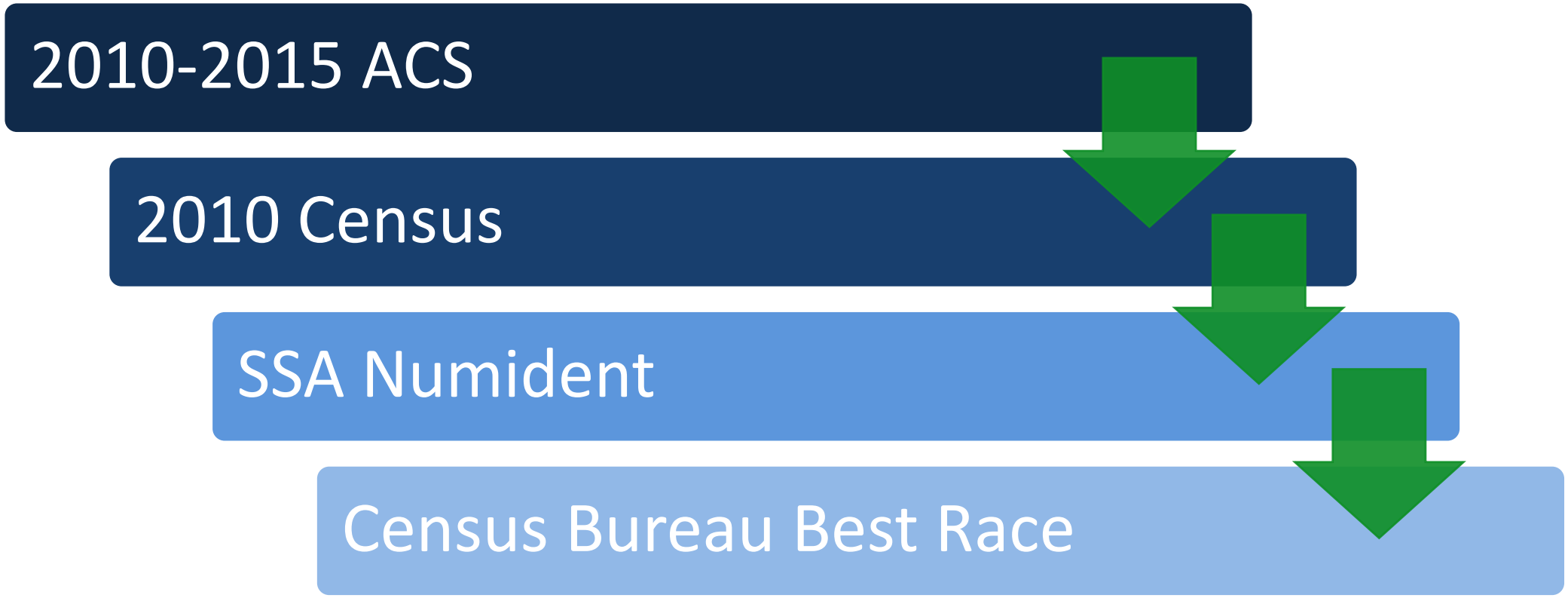
United States™ Census Bureau

Case Study: Integrating AR in the ACS Edit and Imputation Procedures

- Uses 2016 ACS response data
- Test items: Age, Sex, Race, Hispanic Origin, Place of Birth
- Incorporates survey, census, and AR data in lieu of imputation
- Simulated data will be run through ACS edits to create estimates to compare with published 2016 ACS estimates
- Preliminary research:
  - Does not include full edit run
  - Cannot determine full impact of edit run on other survey items or clean-up of AR data
  - Replaces missing with AR
  - Uses that file to determine preliminary estimates

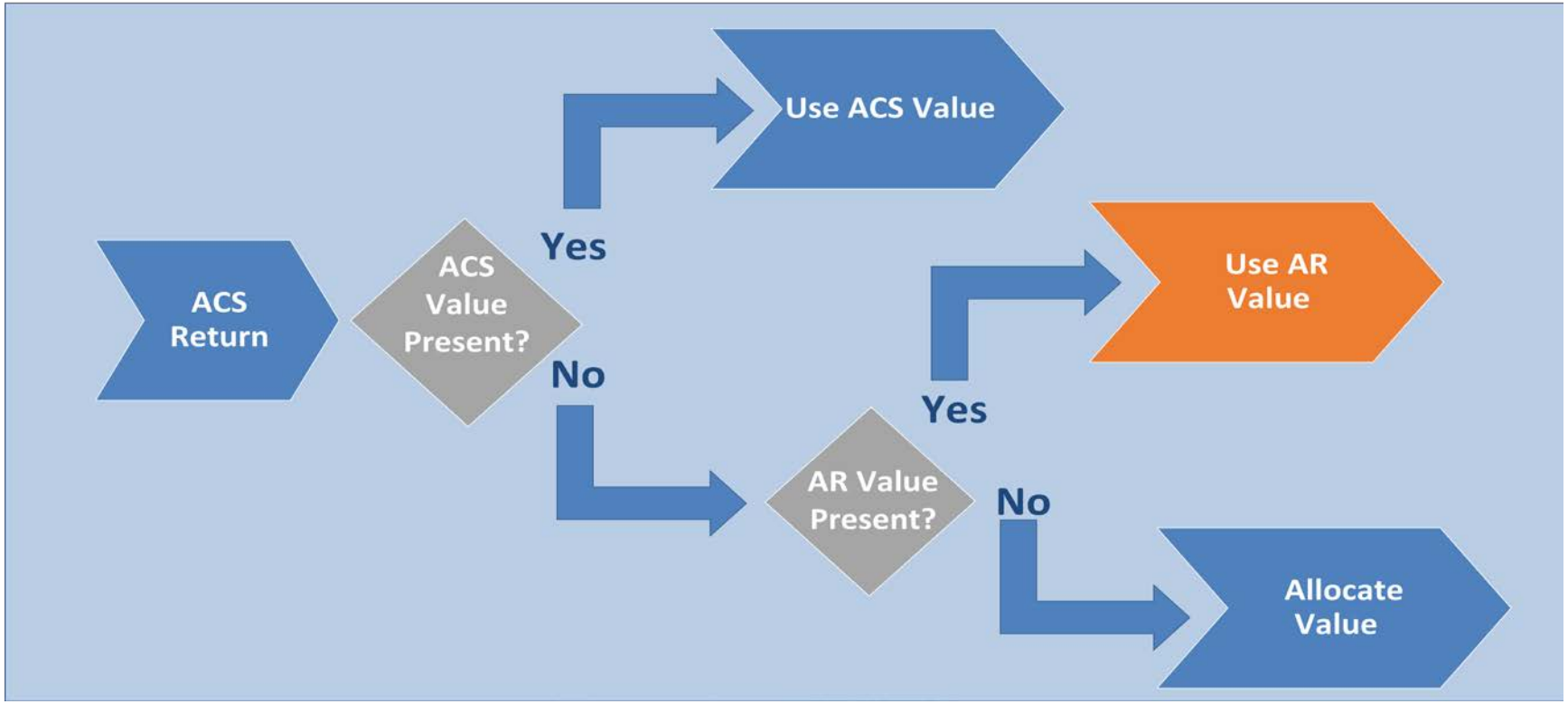
United States™ Census Bureau

Hierarchy of AR Sources



United States Census Bureau

Integrating AR in the ACS Edit and Imputation Procedures – Adaptive Design



United States Census Bureau

Research Questions

1. What proportion of ACS respondents do not provide a response?
2. What percent of missing values can be filled-in with available AR data?
3. What proportion of the AR values match the ACS allocated values?
4. Does using AR change the distribution of the tested items when compared to published estimates?
5. Is there any effect on other survey items, besides those included in the test?

United States Census Bureau

Preliminary Findings

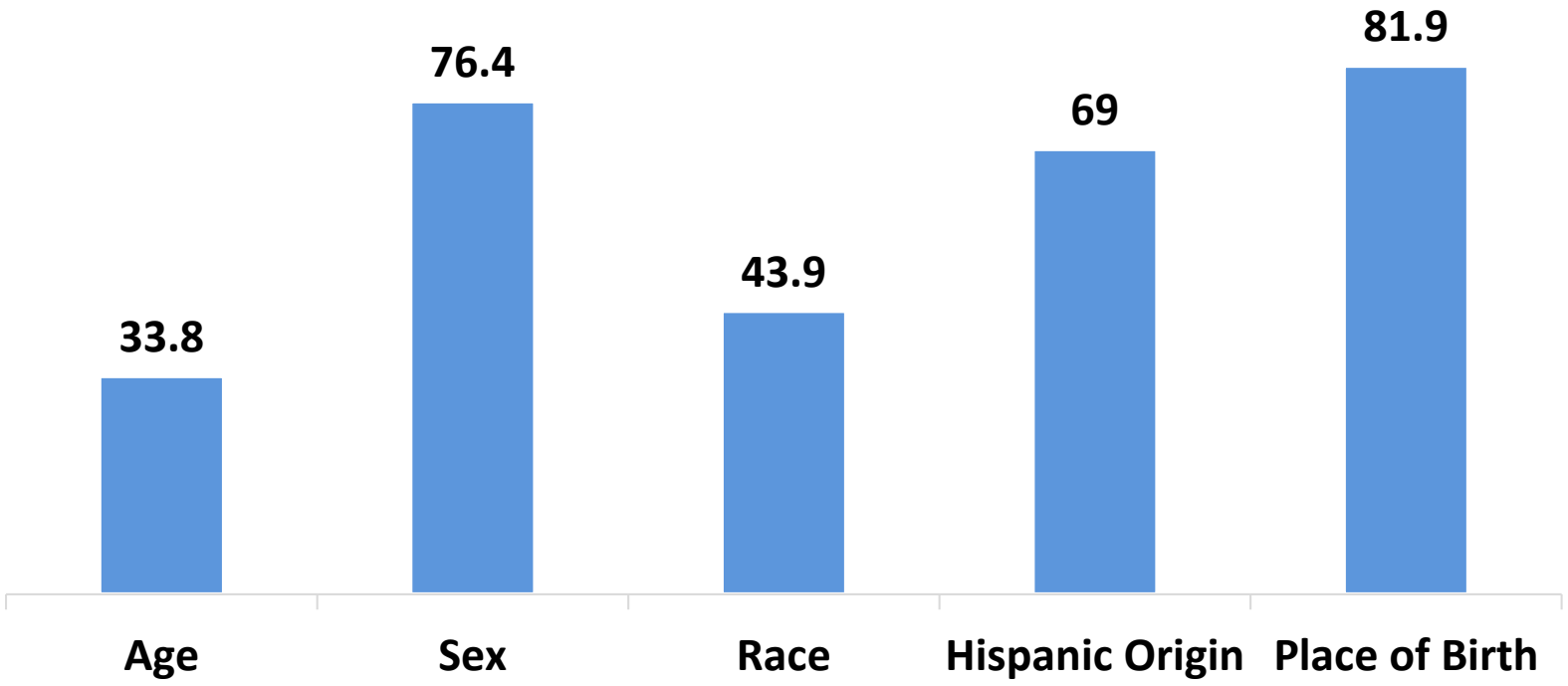
1. What proportion of ACS respondents do not provide a response?

   - Age – **1.0%**
   - Sex – **0.4%**
   - Race – **1.6%**
   - Hispanic Origin – **1.6%**
   - Place of Birth – **6.8%**

United States Census Bureau

Preliminary Findings

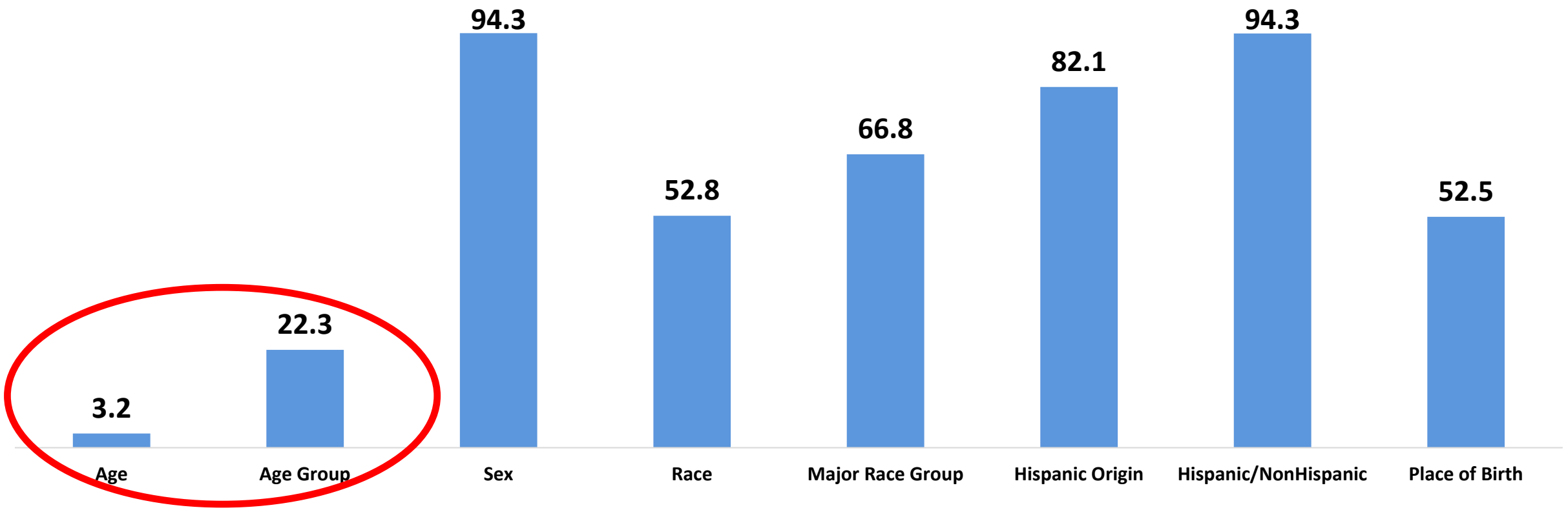
2. What percent of missing values can be filled-in with available AR data?



United States Census Bureau

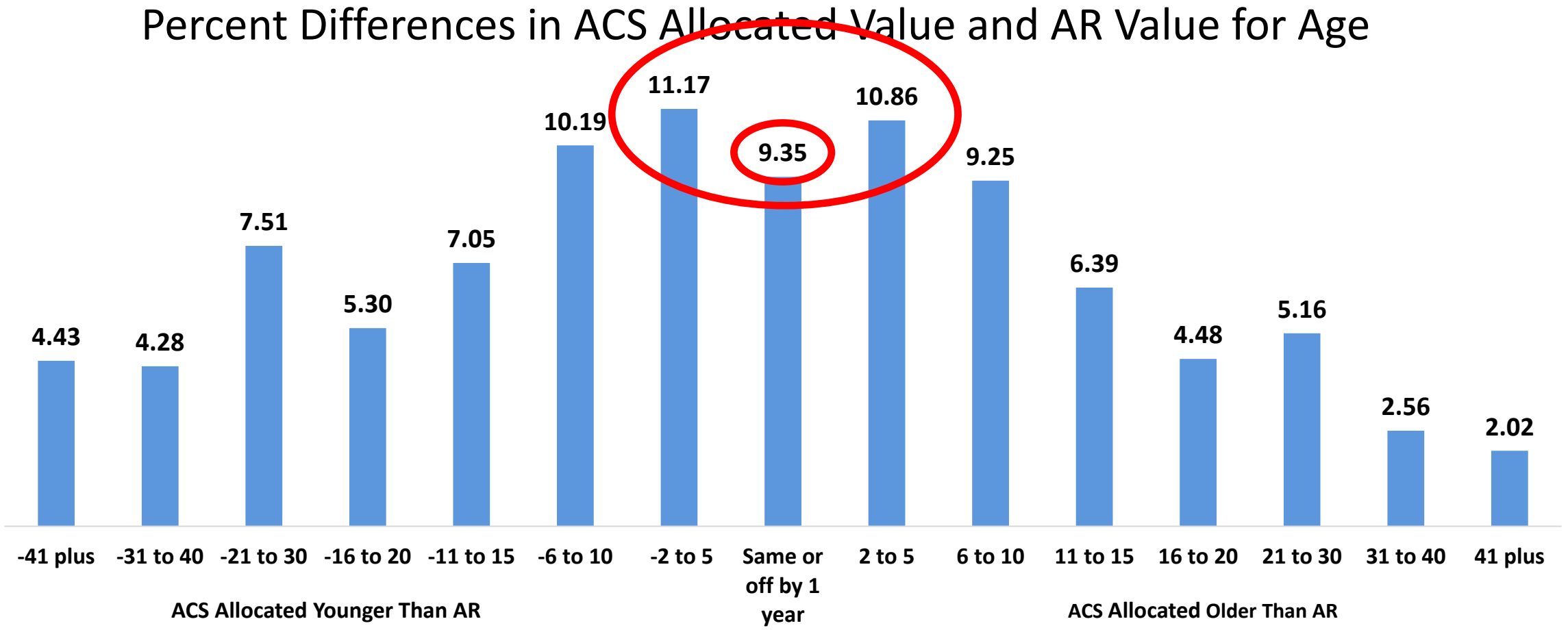
Preliminary Findings

3. What proportion of the AR values match the ACS allocated values?



United States Census Bureau

Preliminary Findings



United States Census Bureau

Preliminary Conclusions

- AR is available to fill-in a large percent of missing ACS values
- Fairly low match rates suggests ACS edits may not accurately capture missing response
- Using AR in lieu of statistical approaches may improve data quality

United States Census Bureau

Next Steps

- Study the differences between AR values and ACS allocated values
- Run the test dataset with AR through ACS edits and compare resulting item distributions with published distributions
- Determine if using AR to fill-in for missing values impacts other items not included in the test
- Develop method to quickly apply edits to test additional items for AR allocation
  - Using AR to allocate values for other items with higher missing data rates will provide larger benefit to ACS program

United States™ Census Bureau

Using Administrative Data in ACS Production – Tentative Target Dates

	2016-2020	2021	2022	2023	2025	2030
Evaluating Administrative Sources						
Demographic Items		Age, Sex, Race, Hispanic Origin	Place of birth, Citizenship			
Housing Items			Acreage	Year Built, Property Taxes, Property Value		
Income Items					Question Changes Reference Period	Applicable Income Items

United States Census Bureau

Linking ACS and IRS Data to Assess Educational Attainment by Family Income

Leah Clark
Center for Economic Studies

United States Census Bureau

Tracking educational attainment by childhood family income

- Cross-sectional surveys like the American Community Survey (ACS) offer measures of educational attainment on a national scale
  - Pros: Large sample size, collected annually
  - Con: Lacks reliable information about childhood family income
- Studying this topic with other data sources brings up other challenges
  - Longitudinal surveys have information about childhood family income but smaller sample size and fewer cohorts
    - For example, Bailey & Dynarski (2011) use National Longitudinal Surveys of Youth data to assess college attendance and completion, but are limited to two cohorts
  - Administrative records offer annual population-level statistics, but are limited in scope
    - For example, Chetty et al. (2014) use tax records to study college attendance, but cannot assess college completion

United States® Census Bureau

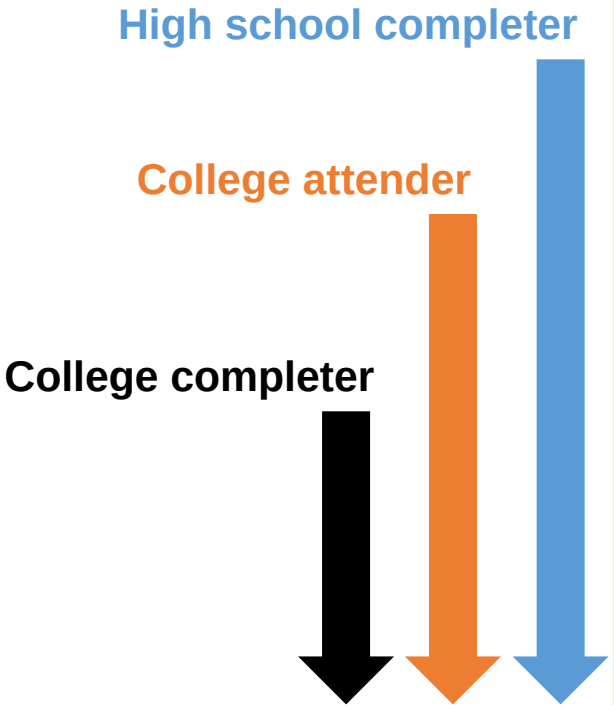
Our contribution

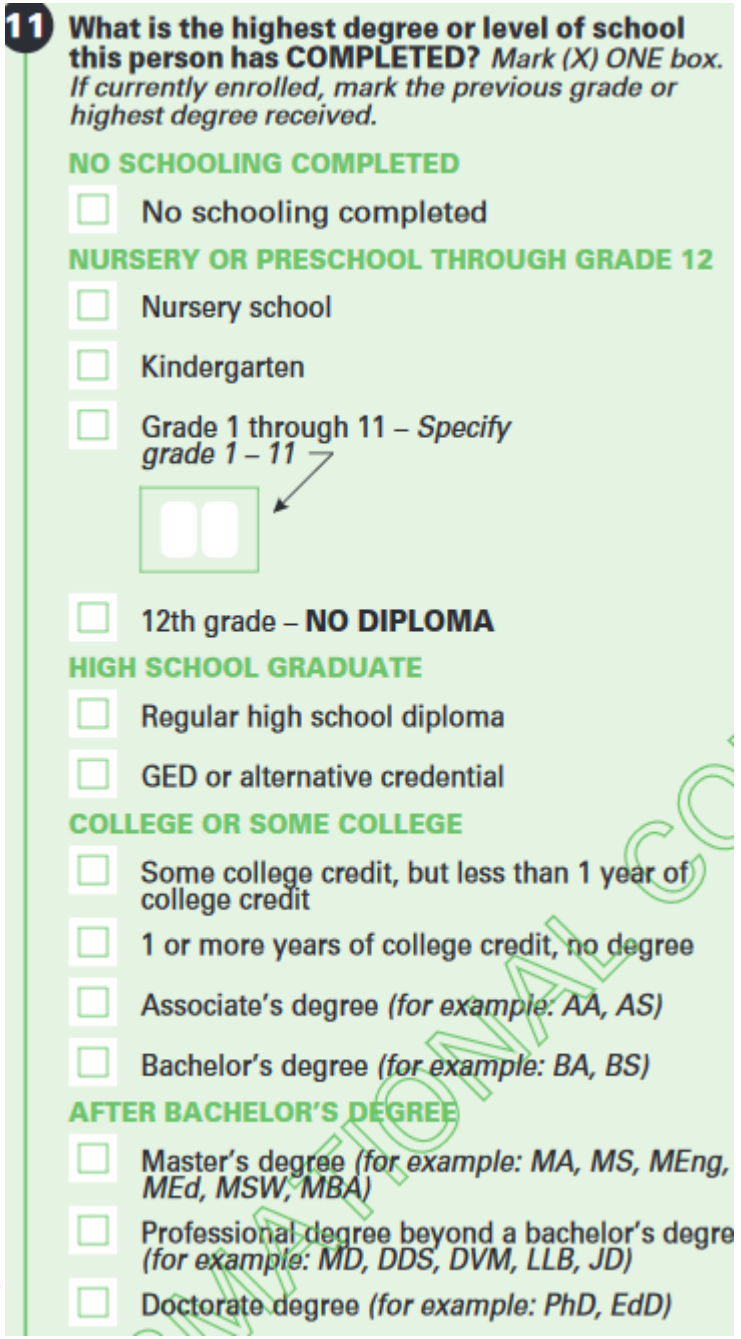
- We link ACS data to Internal Revenue Service (IRS) form 1040 data to bring a measure of childhood family income into the ACS
  - Making a cross-sectional survey longitudinal by leveraging administrative data (Dynarski 2014)
- We produce reliable, annual statistics on high school completion, college attendance, and college completion by childhood family income
  - Cohorts born from 1983-1991
  - Paired with detailed demographic and geographic characteristics, this represents a major step forward for measuring inequality in educational attainment

United States® Census Bureau

Defining educational outcomes

- Derived from ACS question concerning *highest* level of educational attainment
- High school completion includes regular high school diploma, GED/alternative credential attainment or higher
- College attendance includes “some college credit” or higher
- College completion is defined as a bachelor’s degree or higher
  - Measures that include associate’s degrees are also an option




11 **What is the highest degree or level of school this person has COMPLETED?** *Mark (X) ONE box. If currently enrolled, mark the previous grade or highest degree received.*

NO SCHOOLING COMPLETED

- ☐ No schooling completed

NURSERY OR PRESCHOOL THROUGH GRADE 12

- ☐ Nursery school
- ☐ Kindergarten
- ☐ Grade 1 through 11 – *Specify grade 1 – 11*
- ☐ 12th grade – **NO DIPLOMA**

HIGH SCHOOL GRADUATE

- ☐ Regular high school diploma
- ☐ GED or alternative credential

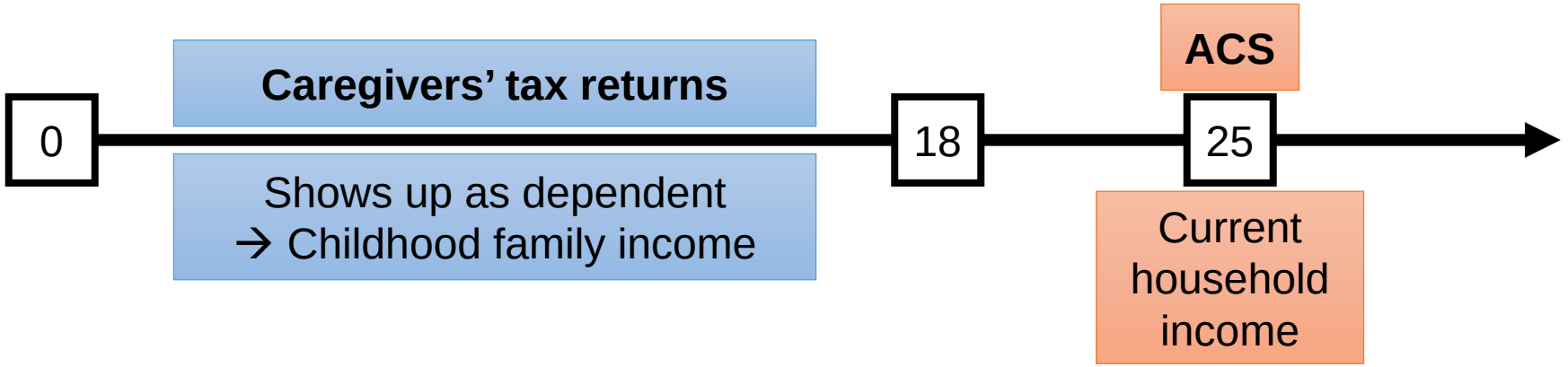
COLLEGE OR SOME COLLEGE

- ☐ Some college credit, but less than 1 year of college credit
- ☐ 1 or more years of college credit, no degree
- ☐ Associate’s degree *(for example: AA, AS)*
- ☐ Bachelor’s degree *(for example: BA, BS)*

AFTER BACHELOR’S DEGREE

- ☐ Master’s degree *(for example: MA, MS, MEng, MEd, MSW, MBA)*
- ☐ Professional degree beyond a bachelor’s degree *(for example: MD, DDS, DVM, LLB, JD)*
- ☐ Doctorate degree *(for example: PhD, EdD)*

Defining childhood family income



1. Locate ACS respondents as dependents in IRS form 1040 data
2. Pull household adjusted gross income during the years they turned 15, 16, or 17, and adjust for inflation
3. Define childhood family income as the average of those three years, ignoring missing or negative values
4. Rank respondents by childhood family income within each birth cohort and split into three equal-sized groups (high, middle, low)

United States® Census Bureau

New linked dataset

ACS 2006-2017

- Respondents born 1983-1991
  - Surveyed at ages 24-26
- Variables:
  - Educational attainment
  - Race/ethnicity
  - Sex
- Excludes noncitizens

+

Administrative Records

- IRS form 1040
  - 1998-2014
  - Variable: Childhood family income
- Census Numident
  - Social Security Administration records
  - Variable: Date of birth

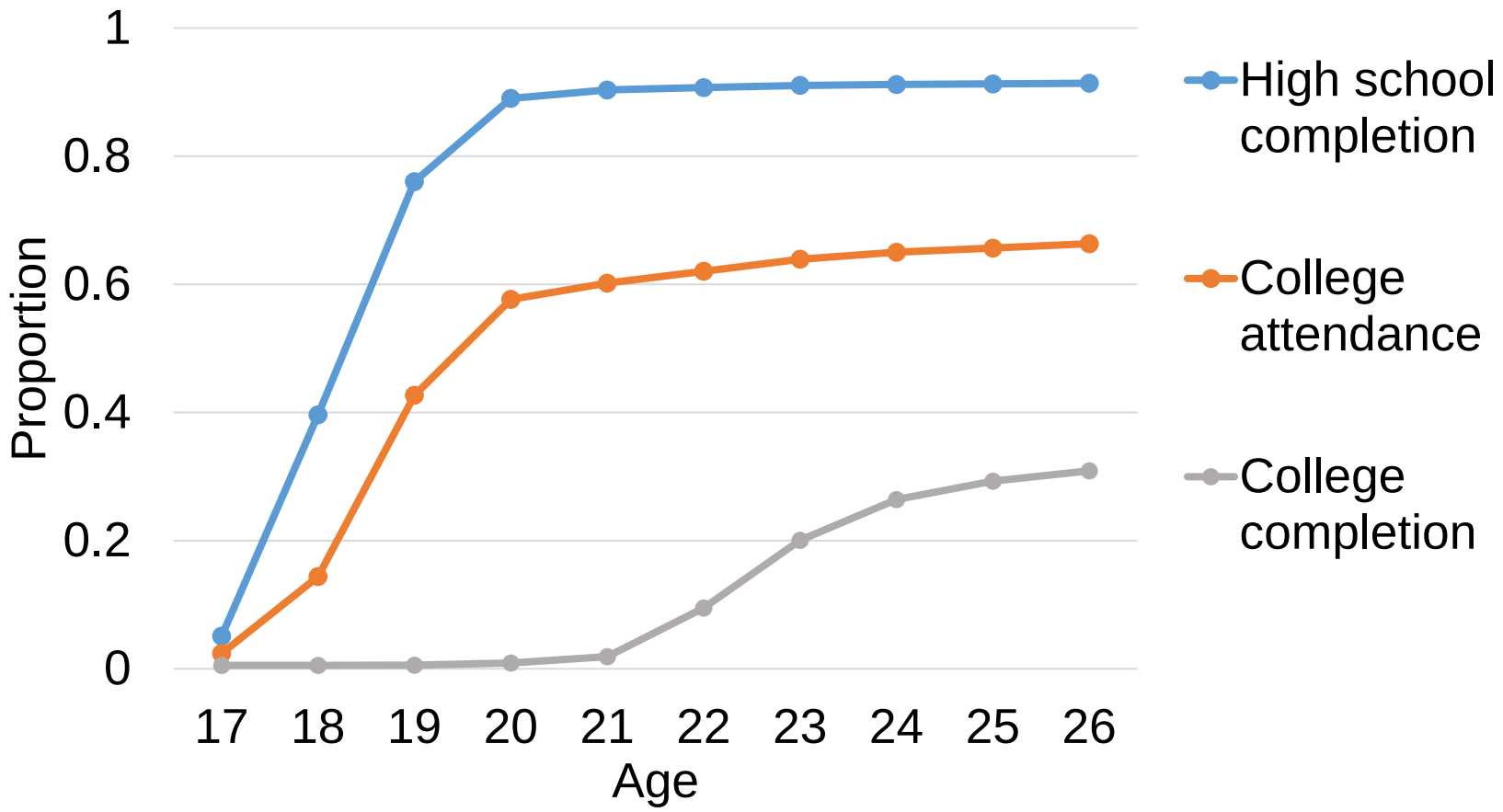
=

Over 1 million linked respondents

- Childhood family income measures for 87% of the population
- Sample weights rescaled by inverse probability of having a measure of childhood family income

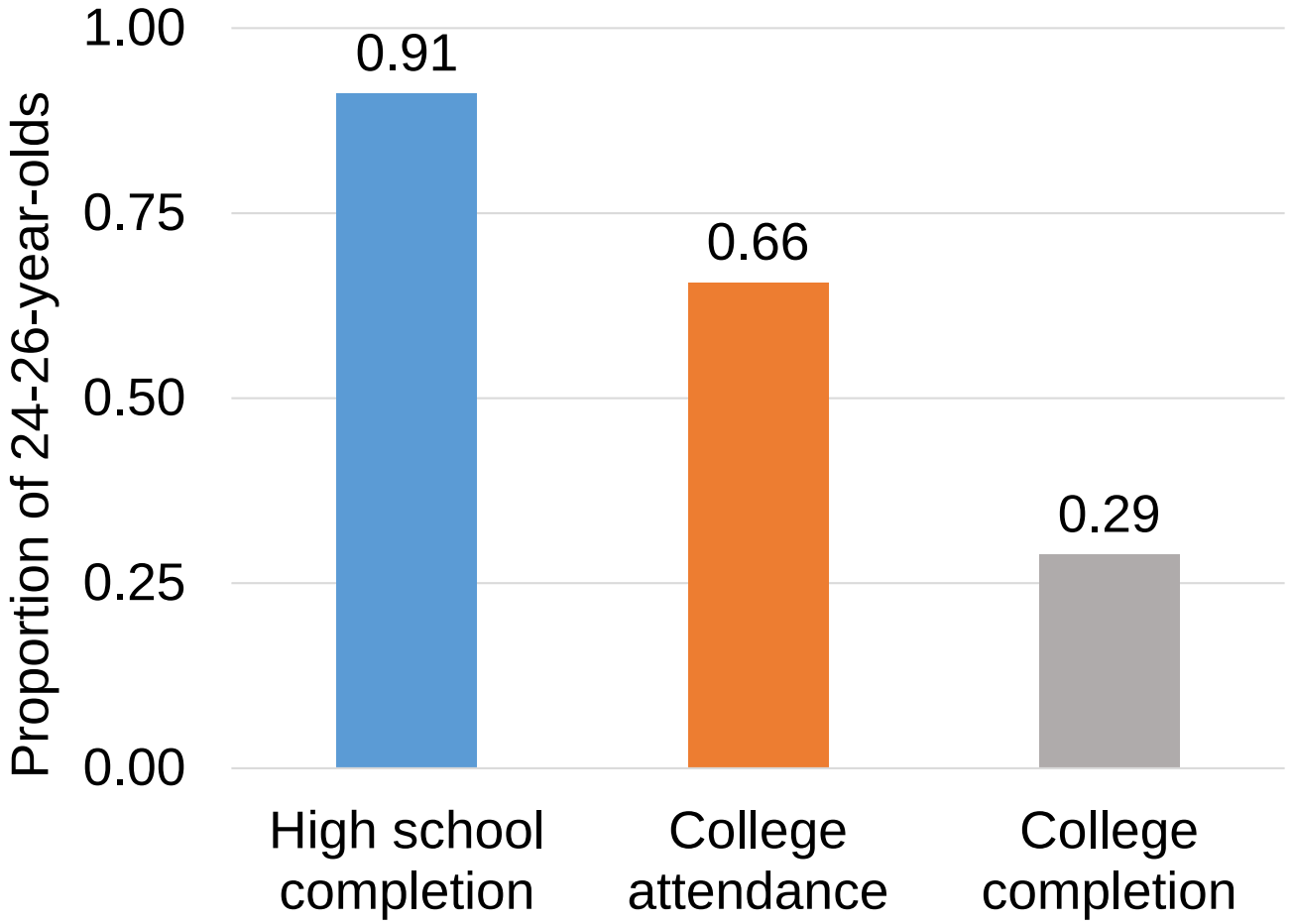
United States® Census Bureau

Educational attainment by age



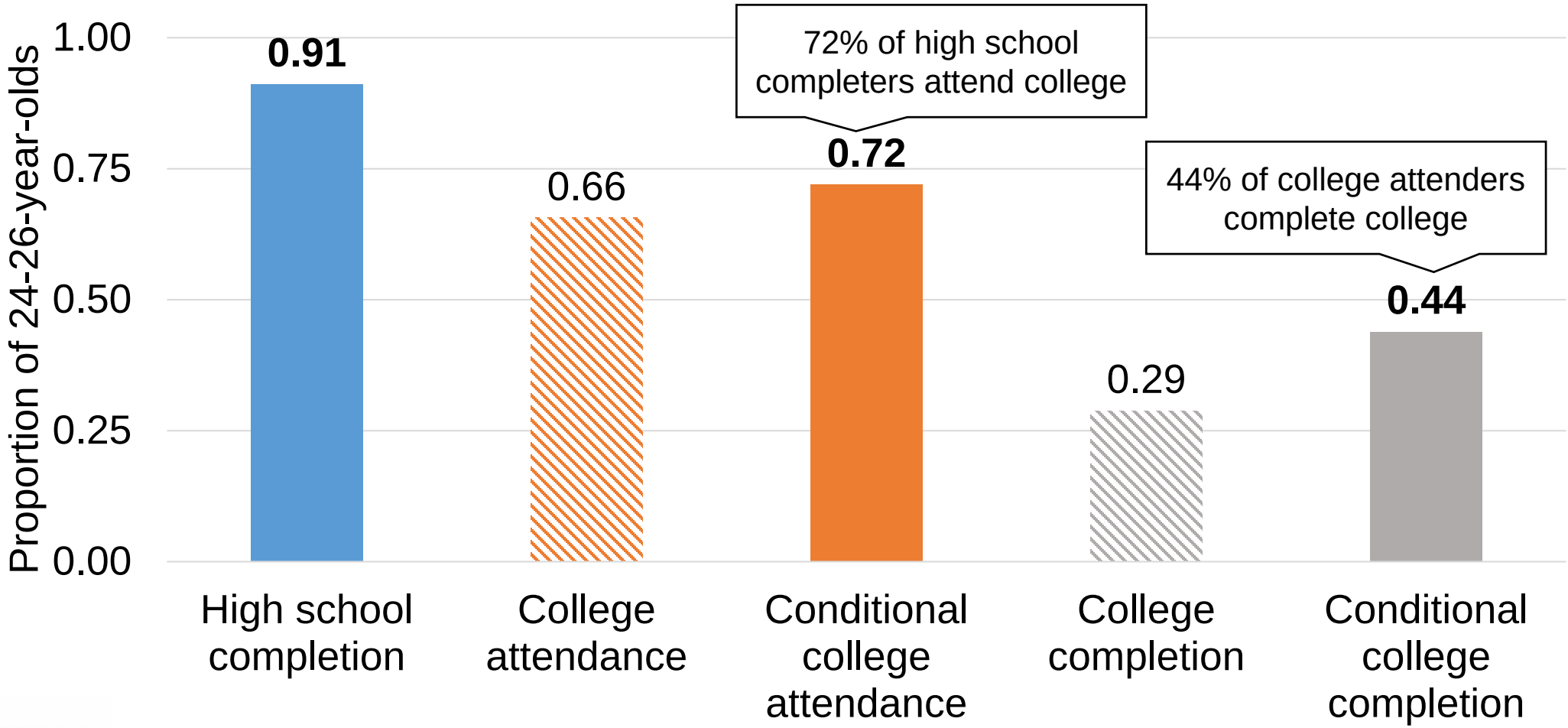
Source: ACS 2006-2017, Census Numident and IRS 1040s 1998-2014

What is the educational attainment of young people born between 1983 and 1991?



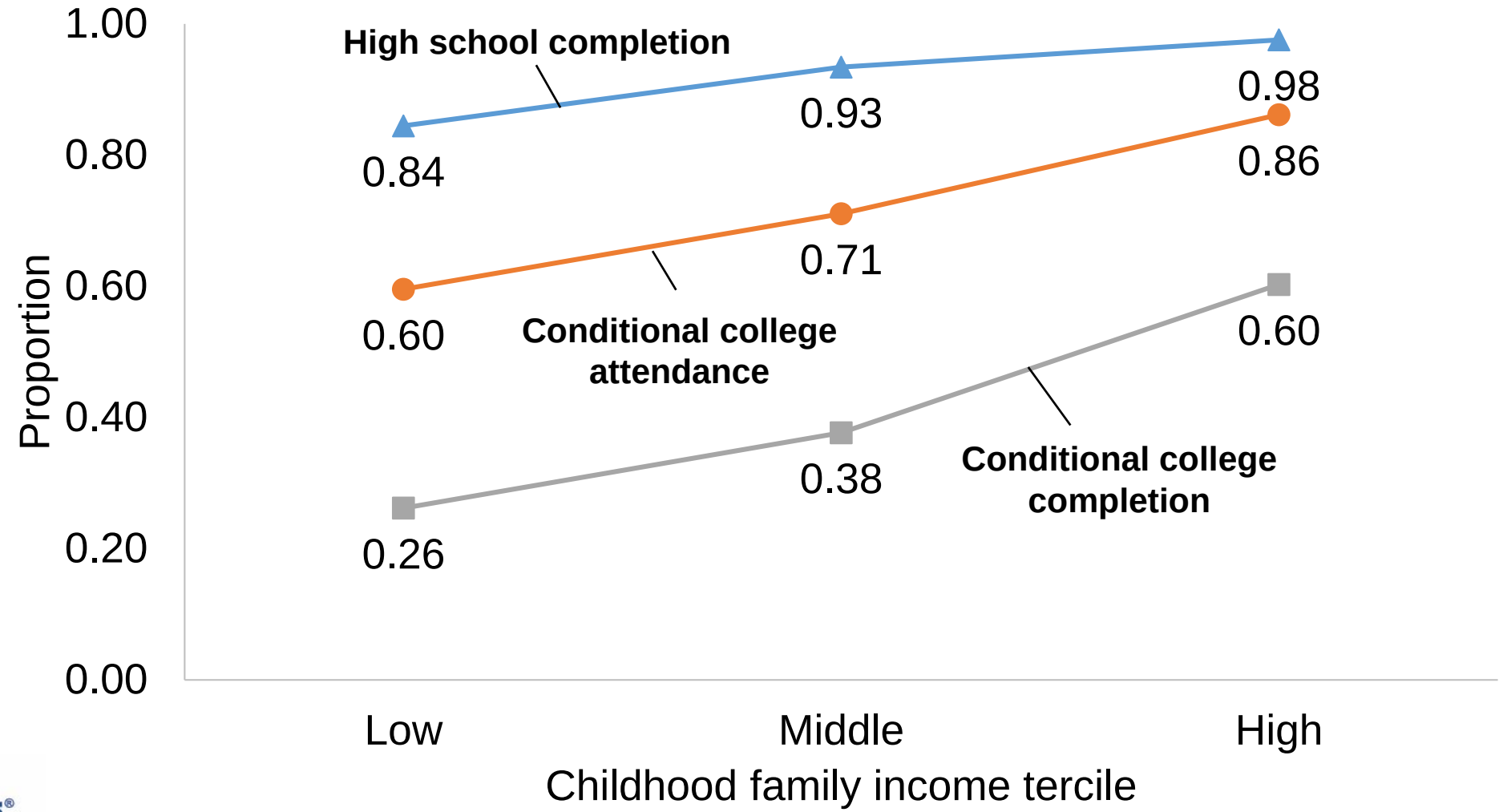
Source: ACS 2006-2017, Census Numident and IRS 1040s 1998-2014

Conditional rates allow us to isolate the inequality that arises at each education level



Source: ACS 2006-2017, Census Numident and IRS 1040s 1998-2014

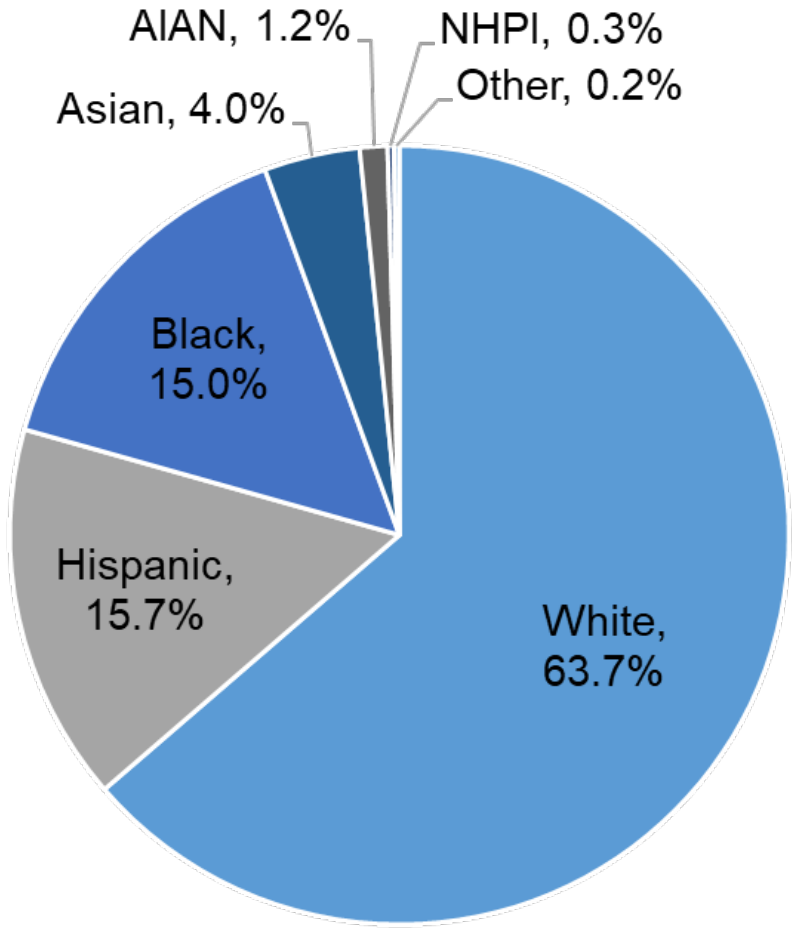
How does educational attainment vary by family income?



Source: ACS 2006-2017, Census Numident and IRS 1040s 1998-2014

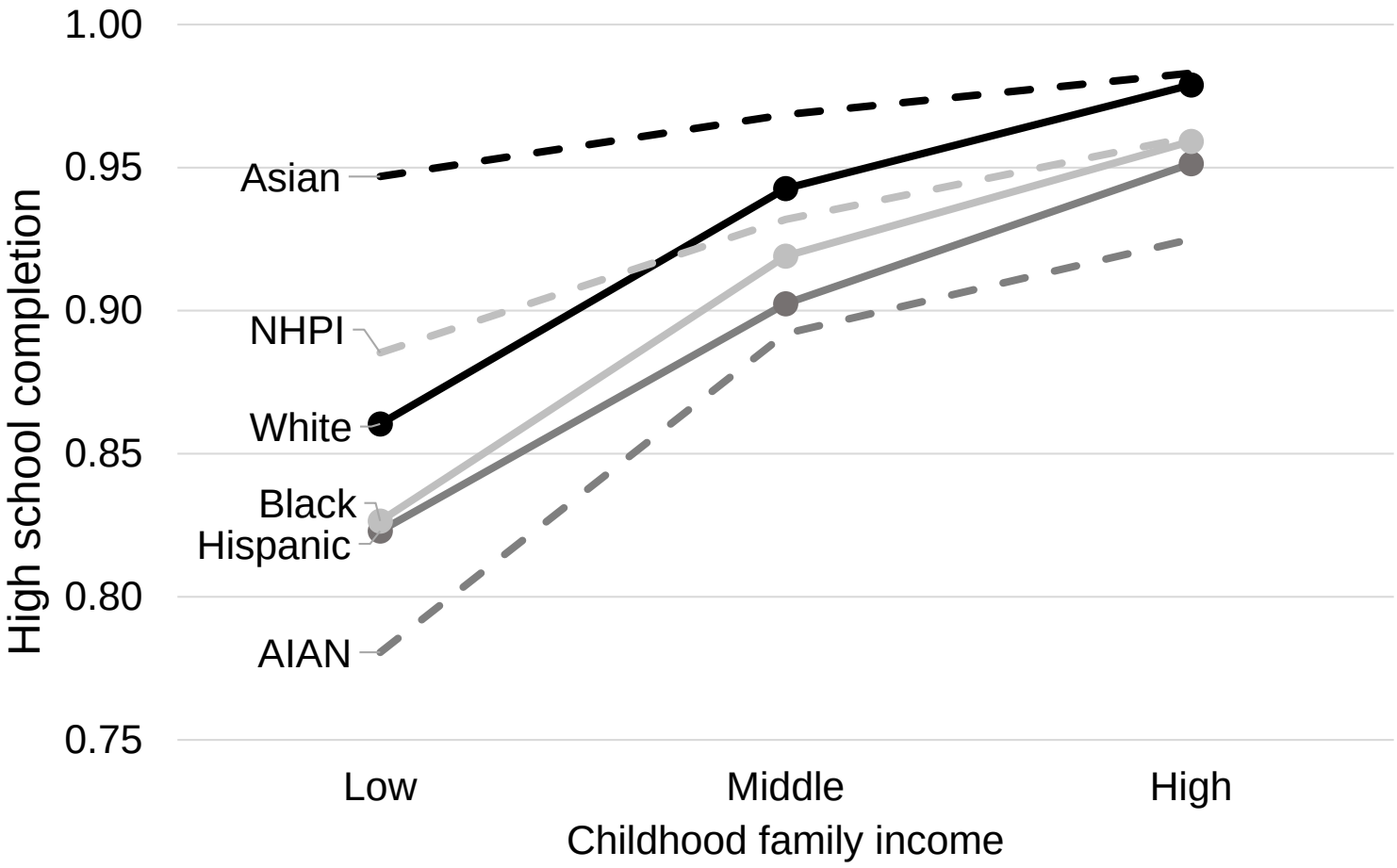
Race/ethnicity

- We use the following race/ethnicity categories:
  - Hispanic (of any race), and
  - Non-Hispanic:
    - White
    - Black
    - Asian
    - American Indian/Alaskan Native (AIAN)
    - Native Hawaiian/Pacific Islander (NHPI)
    - Other



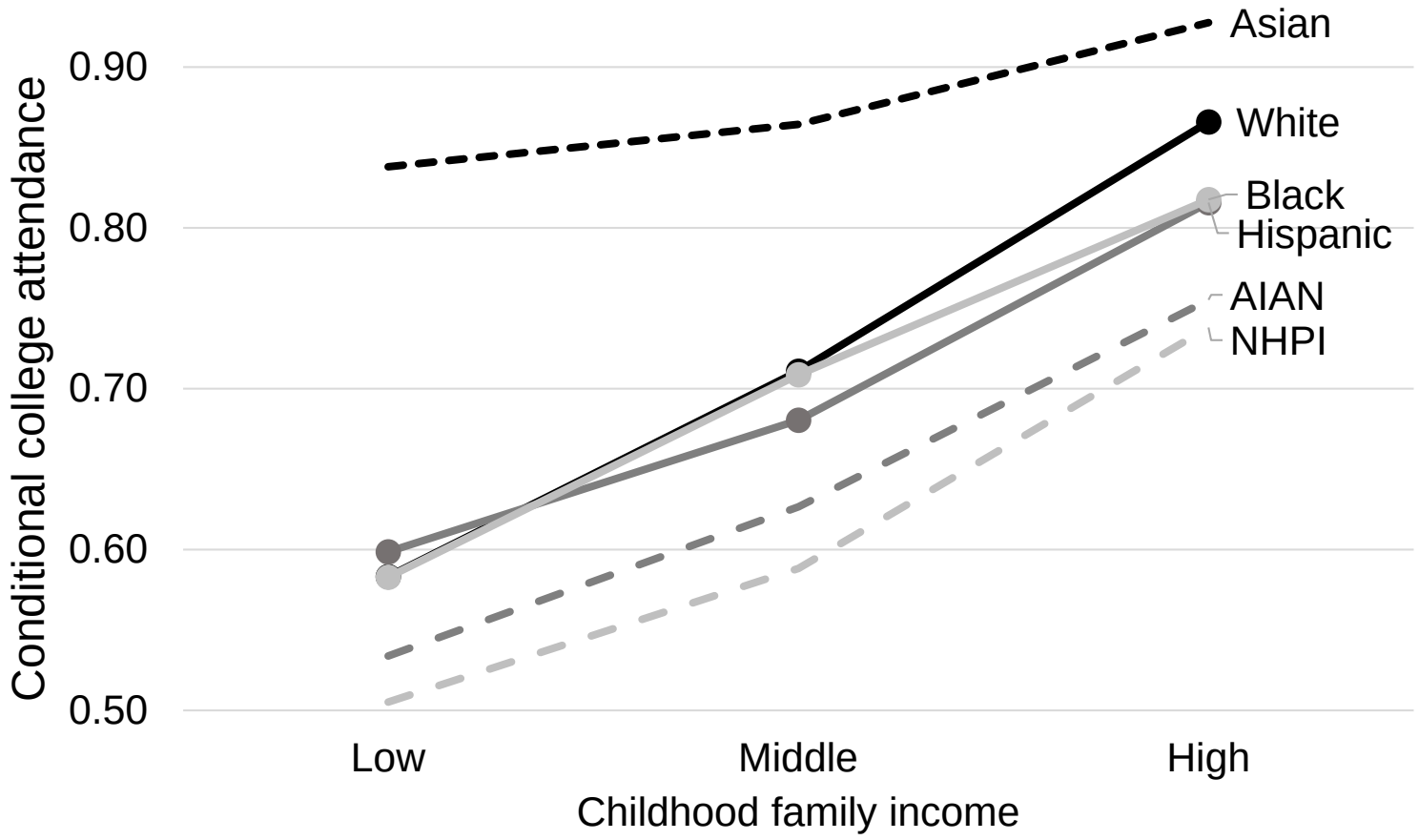
Source: ACS 2006-2017, Census Numident and IRS 1040s 1998-2014

High school completion by income & race



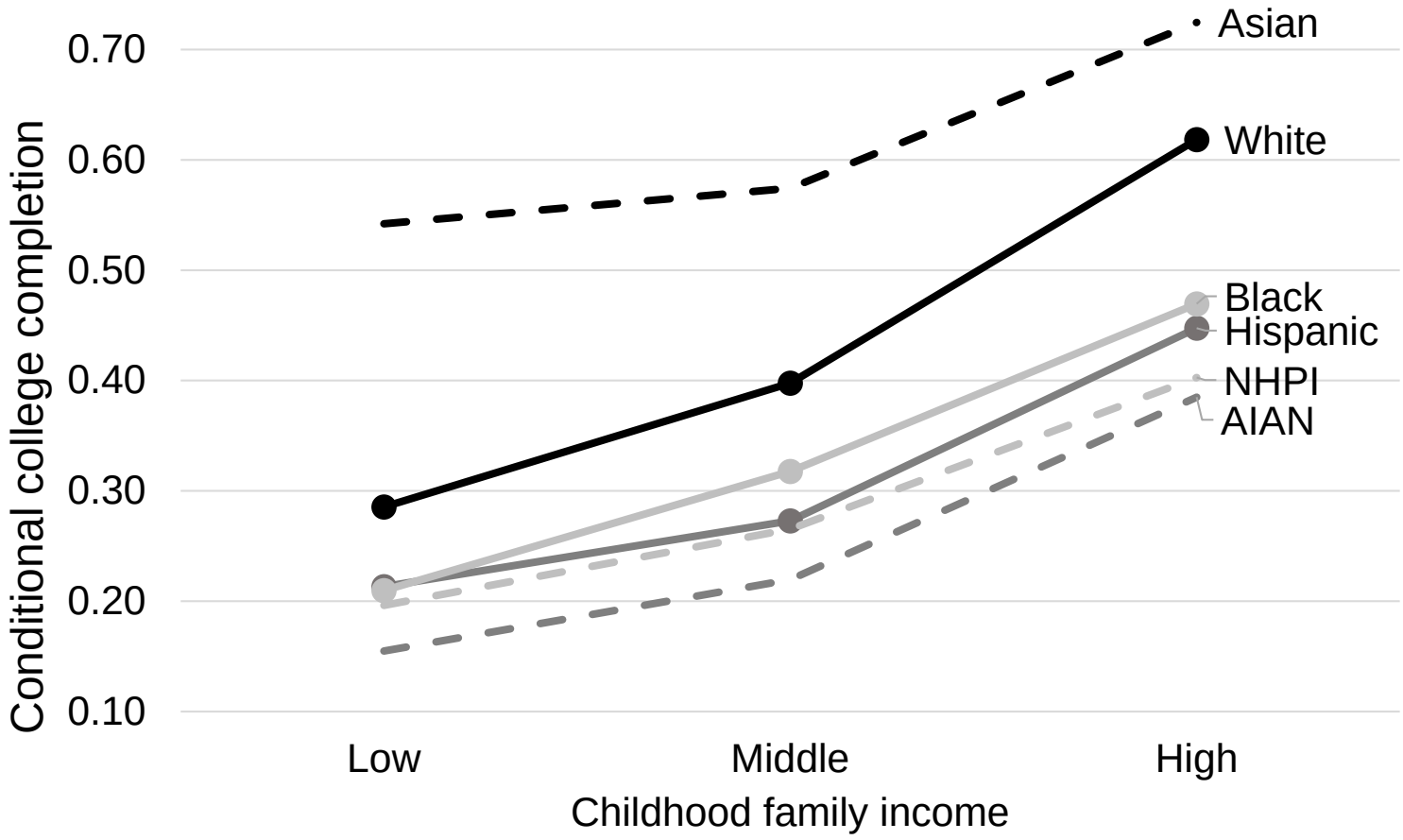
Source: ACS 2006-2017, Census Numident and IRS 1040s 1998-2014

Conditional college attendance by income & race



Source: ACS 2006-2017, Census Numident and IRS 1040s 1998-2014

Conditional college completion by income & race

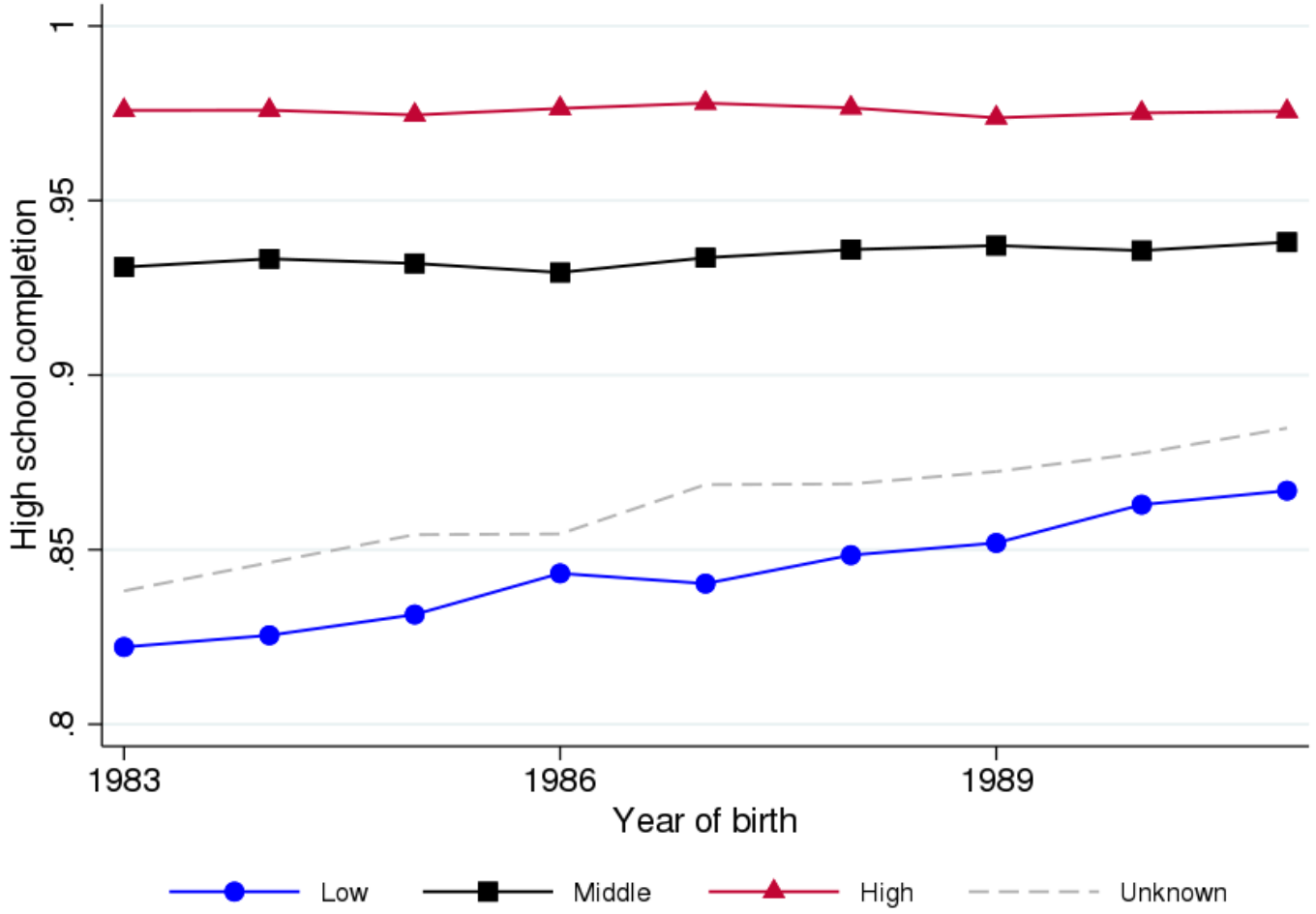


Source: ACS 2006-2017, Census Numident and IRS 1040s 1998-2014

Trends in educational attainment

- Rather than pool over birth years, we can plot attainment for each year and income group
- We use regressions to test changes in income gaps
  - Can also test changes by income and race

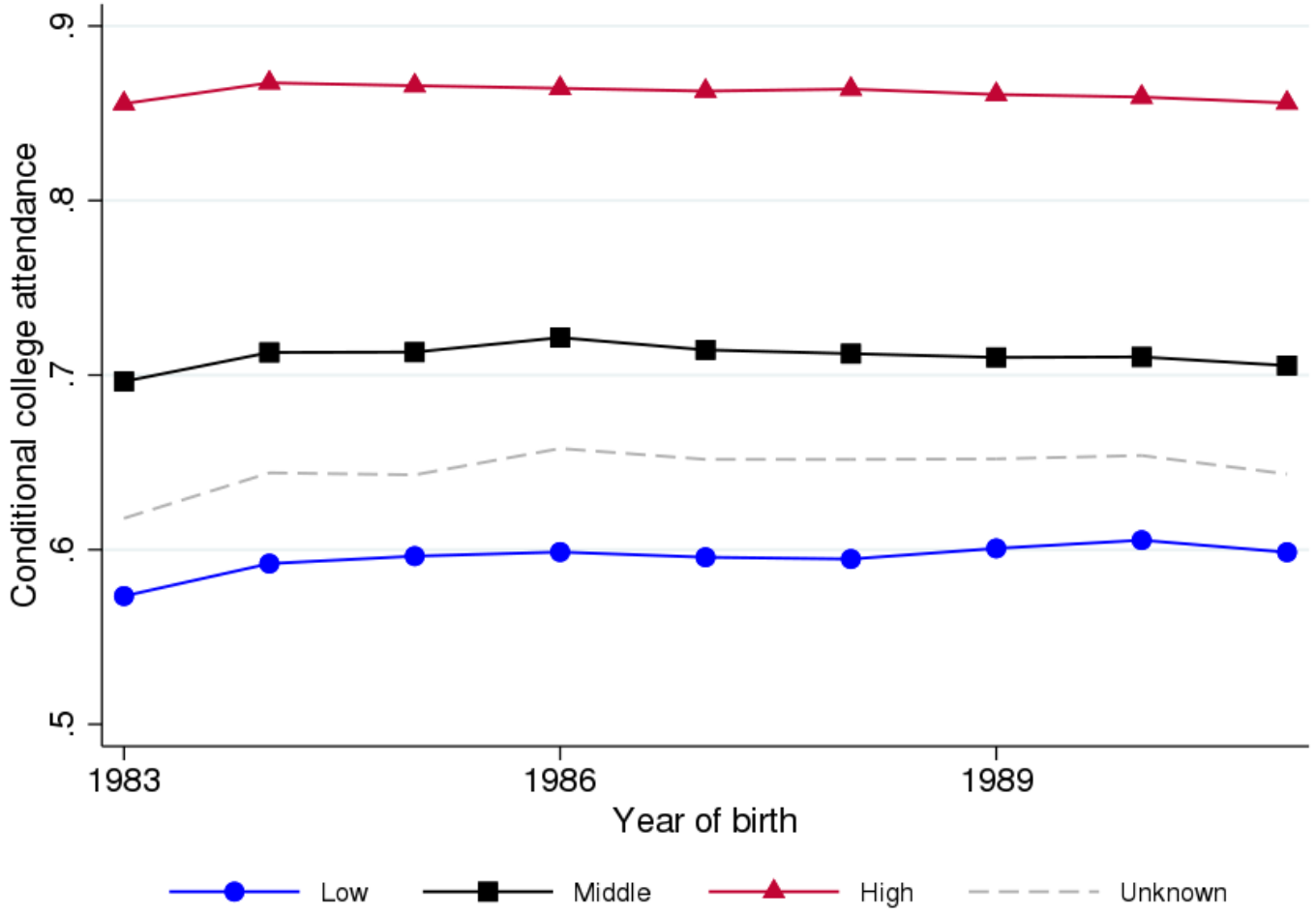
Low-income young people made gains in high school completion



- High school completion rate for low-income young people grew by 4 percentage points
- The income gap in high school completion narrowed by more than one-fourth

Source: ACS 2006-2017, Census Numident and IRS 1040s 1998-2014

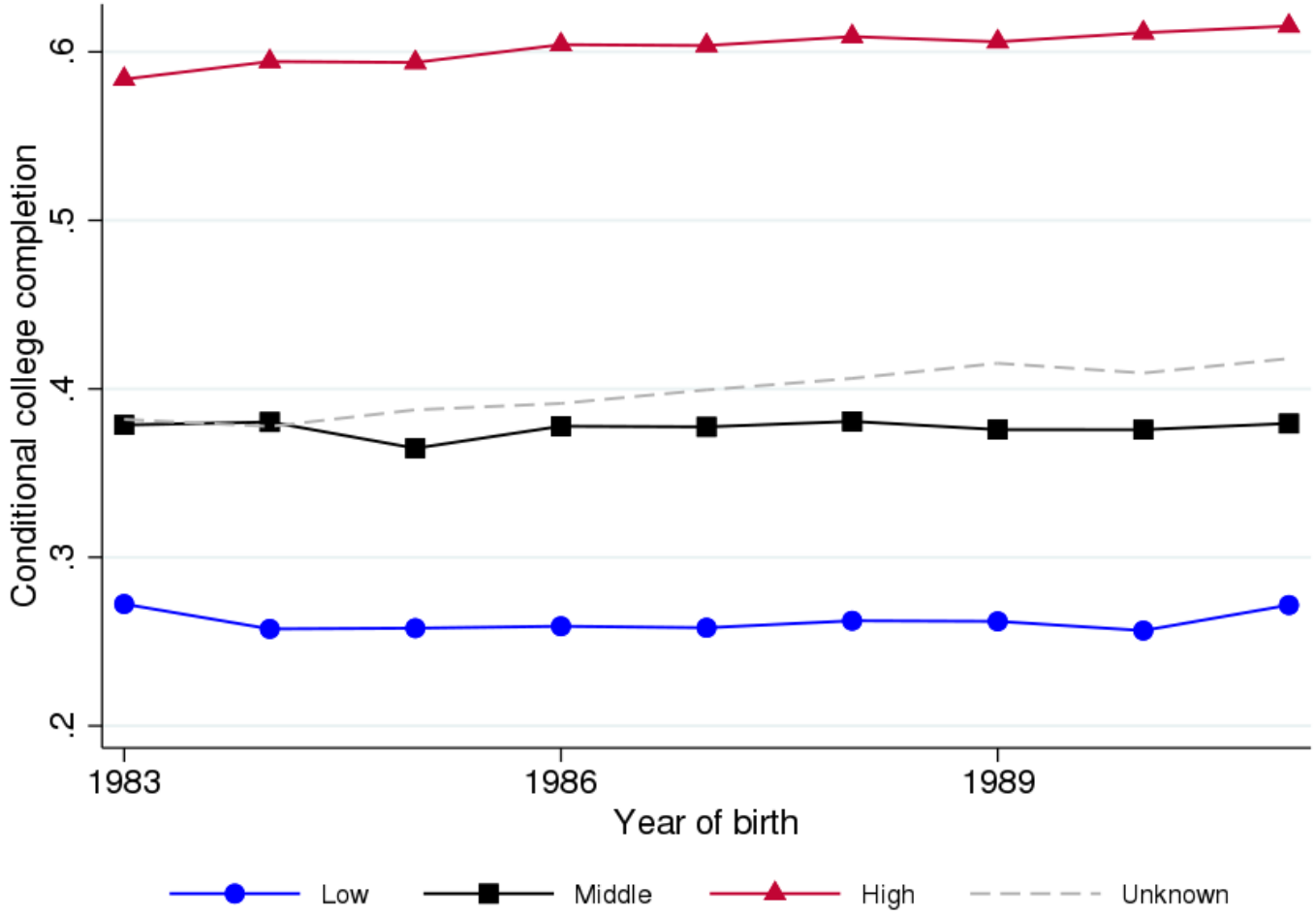
Small low-income gains in college attendance



- Low-income high school completers made very modest gains in college attendance
- The income gap narrowed by less than one-tenth

Source: ACS 2006-2017, Census Numident and IRS 1040s 1998-2014

The income gap widened in college completion



- High-income college attenders became more likely to complete college
- The income gap continued to grow

Source: ACS 2006-2017, Census Numident and IRS 1040s 1998-2014

Trends expand on existing literature

- High school graduation rate stagnant for young people born before 1980, but rising for cohorts born in the 1980s, especially for Black and Hispanic young people (Heckman & LaFontaine 2010; Murnane 2013)
  - We document that rise is driven by low-income young people
  - Low-income Black and Hispanic respondents make large gains, but low-income AIAN respondents do not make gains
- The income gap in college attendance rose for cohorts born 1961-1982 (Bailey & Dynarski 2011)
  - Like Chetty et al. (2014), we find a slight decline for cohorts born in the 1980s
- Bailey & Dynarski also find growth in the income gap in college completion
  - We show that the gap has continued to grow

United States® Census Bureau

Conclusion

- Linking measures from administrative data to cross-sectional surveys opens many new analysis opportunities
  - Cross-generational income measure
- Large and persistent disparities in educational attainment by childhood family income
  - Low-income gains in high school completion
- Income gaps vary across racial subgroups, but they are consistently large within subgroups
- Rates of educational attainment differ by racial subgroups, and so do trends
  - American Indian and Alaskan Native young people experience low levels of educational attainment, and are not showing signs of convergence with other subgroups

United States® Census Bureau

The American Community Survey

Data Tell Stories. Tell us Yours!

Your story

Your data details

Your outcome

census.gov/acs/www/share-your-story

United States® Census Bureau

The American Community Survey

Continue the Conversation

Sign up for and manage alerts at
https://public.govdelivery.com/accounts/USCENSUS/subscriber/new

More information online:
census.gov/acs

Census Customer Service Center
800-923-8282

Social media: @uscensusbureau
#ACSdata

acso.users.support@census.gov

Source Us:

U.S. Census Bureau's [YYYY – YYYY] American Community Survey
[1/3/5]-year [estimates/statistics/data release]

United States® Census Bureau

For More Information

Realizing the Promise of Administrative Data for Enhancing the American Community Survey
https://www.census.gov/programs-surveys/acs/operations-and-administration/agility-in-action/administrative-records-in-the-american-community-survey.html

Housing Administrative Record Simulation
https://www.census.gov/library/working-papers/2018/acs/2018_Clark_01.html

Housing Administrative Record Simulation Data Visualization
https://www.census.gov/library/visualizations/interactive/housing-admin-record-simulation.htm

Reports Evaluating Administrative Data Sources
https://www.census.gov/content/census/en/programs-surveys/acs/library/publications-and-working-papers/research-and-evaluation.All.html/

Contact the Presenters:

Andrew Keller	Andrew.D.Keller@census.gov
R. Chase Sawyer	Robert.C.Sawyer@census.gov
Sandra L. Clark	Sandra.L.Clark@census.gov
Leah Clark	Leah.R.Clark@census.gov

United States Census Bureau